DNA Barcodes and Meiofaunal Identification

Jenna D. Mann

Thesis submitted in accordance with requirements for the degree of

Doctor of Philosophy

The University of Edinburgh

2009

For Granny

Image on title page: Tardigrade (specimen CA10) ingesting a nematode

Declaration

I declare that this thesis has been composed by me and is the result of my own work. Work done in collaboration is explicitly stated.

J. D. Mann

October 2009

Abstract

In recent years there has been a desire to definitively catalogue the life on our planet. In light of the increasing extinction rates that are driven by human activities, it is unlikely that this will be achieved using traditional methods. Whilst most organisms which have a body size of more than 1cm have been described, the vast majority of animal life is smaller than this, collectively known as meiofauna, and is yet to be catalogued. Meiofaunal organisms present a range of problems for traditional taxonomy. Firstly they are microscopic, meaning that morphological features are often difficult to resolve. Secondly these creatures often exhibit cryptic diversity meaning that different species often look the same. Thirdly, it is often the case that the organisms are poorly described in the literature making it very difficult to confirm identification, assuming that someone has already described it. It is possible, however, to obtain DNA sequences from these organisms. DNA barcoding, the use of short sequences of DNA to identify individuals, is now commonly used in a wide range of applications. It has been proposed that a single target gene should be sufficient to describe all organisms this way. Barcodes can be acquired from individuals or from bulk extractions from environmental samples. In the latter case, many of the sequences obtained are novel and unlikely to ever have a type specimen associated with them. When this is the case, assessing the diversity of a sample becomes a computational exercise. However, as yet, there is no agreed standard method adopted for analyzing the barcodes produced. Indeed most methods currently employed lack objectivity. This thesis investigates the efficiency of a range of gene targets and analysis methods for DNA barcoding, with an emphasis on meiofaunal organisms (nematodes, tardigrades and thrips). DNA barcodes were generated for up to three genes for each specimen. Sequences for each gene were analysed using two programs, MOTU_define.pl and DOTUR. These programs use different methods to assign sequences to operational taxonomic units (OTU), which were then compared. An objective method for analysing sequences such as MOTU_define.pl, which relies on only the information contained in the sequences, was found to be most suitable for designating taxa. It does not

attempt to apply evolutionary models to the data, and then infer taxa from the derived data. In addition to barcoding, some samples were pre-processed using video capture and editing (VCE). This creates a virtual slide of a specimen so that a sequence can be linked to a morphological identification. VCE proved to be an efficient method to preserve morphological data from specimens.

Acknowledgements

I have many, many people to thank for helping me through this PhD.

Firstly I'd like to thank everyone who helped with training, providing samples and furthering my scientific skills. Dr Sue Hockland, Tom Prior, James Quill and Bex Lawson (CSL, York) to introducing me to Aphelench nematodes and Dr Tom Bongers and Hanny van Megen (Wageningen University, Netherlands) for unveiling the rest of the nematodes to me. I also need to thank all those who have generously taken the time to teach me lab manners and techniques, Fran Thomas and Claire Conlon who taught me during my undergraduate project and Marian Thomson. I'd also like to thank members of Dr Paul de Ley's lab (UCR, USA) for an enjoyable visit to their lab and Melissa Yoder for showing me VCE. Thanks to Stephen Lewis, Manuel Mota, Ann Burnell and CSL for providing nematodes for my work and Dr Dom Collins and Rachel Glover (CSL) for providing the thrips data. I also need to mention Sarah Mayer (Chief Leader) and Susie Allison of the BSES Greenland 2007 expedition for allowing me to tag along and collect samples from Disko Island, and to Jess Abbott who carried them back. Back in Edinburgh, I greatly appreciate all the help of the Gene Pool (Jill Lovell, Andy Gillies, Anna Montazam and Jenna Nichols) who did all my sequencing. And to Sujai Kumar who has written many a perl script to make my life easier.

Secondly, major thanks to all who offered support, moral, physical or otherwise. My parents (who have sent more than one food parcel), my girls for keeping me endless amused with their ratty ways and many friends for providing the odd pint (or two). And to Graham Stone and Neil Boonham, my second supervisors, for all their support and input into this thesis.

Last but not least, thanks to Mark. Without you, I would never have been enamoured by tardigrades or given the opportunity to start this PhD. I appreciate all the time and guidance you've given me, and the wonderful opportunities you've allowed me to take.

Contents

Chapter 1

Introduction to DNA Barcodes and Meiofauna Identification

1.1 A Brief History of Taxonomy

Carl Linnaeus, a mid eighteenth century Swedish botanist, proposed a binomial classification system for plant species. The usefulness of the idea quickly spread and began to be applied to encompass the animal kingdom. At a time when science was the reserve of the rich minority, Linnaeus could not have imagined how universal his system would become. His binomial system combined genus and species identifiers of an organism to give it a unique name. This system had implicit connotations regarding a species taxonomic affinities. Each taxonomic level was a sub-division of a higher classification - species grouped together formed genera which in turned belonged to a particular family of animals.

At the time, species were described and separated on the basis of morphological features. Strict rules were established to describe species thus avoiding both multiple descriptions of the same species (synonymy) and the same name referring to multiple taxa (homonymy) (Godfray, 2002; Minelli, 2003; Tautz *et al.*, 2003). These rules could also include morphological, behavioural and ecological information. Larger animals received most attention from taxonomists, and, more than 250 years later, it is generally accepted that the catalogue of known vertebrate species covers over 90% of total number of species (Waugh, 2007).

The proportion of meiofauna (organisms described as having a body axis of less than 10 mm) which have been described, has been estimated to be less than 50% of the biodiversity of life on Earth (Blaxter, 2003). The comprehensiveness of the larger animals' catalogue may be attributable to the relative ease at which a 'species-concept' can be associated with a particular animal. This is usually based on a variation of the biological species concept whereby a species is said to constitute a population of reproductively compatible individuals (Adams, 1998). However, even for larger animals, a universally accepted species definition has yet to be agreed. In a group of diverse and understudied taxa such as meiofauna, the lack of

an explicit species concept makes species delimitation potentially an inexact science (Adams, 1998; Nadler, 2002).

As we become increasingly aware of how much is still to be described, the call for a complete encyclopedia becomes more persistent (Ronquist and Gardenfors, 2003; Wilson, 2003). As our knowledge of the Earth's biodiversity is growing, we are seeking narrower and more refined species descriptions in an attempt to be ever more precise to meet this need. In addition, as the amount of detail required for a species definition increases, traditional taxonomy struggles under the demands of its prospective workload. There are simply not enough trained taxonomists and enough time to complete the task (Lee, 2000; Tautz *et al.*, 2003; Waugh, 2007; Wilson, 2003)).

Taxonomic expertise is especially wanting for meiofauna, the microscopic animals (e.g. rotifers, tardigrades and nematodes). Recognizing their existence is difficult in the first place. Morphology appears highly conserved (Srivastava *et al.*, 2008) when using light microscopy. Higher resolution techniques reveal astounding diversity of some taxa (De Ley and Blaxter, 2002), and they may well represent the majority of the animal biomass of the Earth (Wilson, 2003). These eukaryotic organisms seem to be ubiquitous and highly speciose. Some taxa, such as nematodes (Phylum Nematoda), have different life stages and can also exhibit phenotypic plasticity (Nadler *et al.*, 2006). Moreover some taxa are asexual and reproduce clonally, making biological species concepts difficult to apply. All these factors make traditional taxonomy and identification incredibly difficult for the non-specialist.

1.2 DNA Taxonomy

In order to catalogue diversity as it is now, the rate at which species are described and identified needs to increase significantly before too much more is lost. For a species description to be accepted, it must appear in print and be subject to rigorous scrutiny by the appropriate peers. A further difficulty is that although such information is available in a large number of publications and in a variety of languages, currently there is no single repository for such information (Gewin, 2002; Minelli, 2003). These factors

make taxonomy problematic, difficult to use, and hinder species description (Godfray, 2002). Centralising and digitization of species descriptions could overcome these issues (Godfray, 2002). By utilizing the Internet, illustrations and other species-related data may be incorporated. Currently, these are not included in all descriptions. Furthermore, by adding historical information about a species' descriptions and synonyms, misidentification may be avoided (Godfray, 2002). That said, the quality of submitted data may also raise concerns (Tautz *et al.*, 2002). Neither is it likely that an increased rate of species descriptions can be achieved solely by increasing the number of taxonomists. It seems this must be accomplished in conjunction with other methods (Tautz *et al.*, 2003). The matter is complex.

A promising new system is DNA taxonomy which uses a set of DNA sequences as a species identifier (Tautz *et al.*, 2003). Under this system, a species description would remain an amalgamation of (morphological, ecological, etc.) information within a species-concept framework, and would have the addition of a DNA sequence (Tautz *et al.*, 2003). Traditional taxonomy requires the nomination of a type specimen as an exemplar for the species. Generating and linking a DNA sequence from this type specimen would provide a firm statement of its molecular definition. Taxonomic definitions can change as new information comes to light and necessitates revision. Renaming of a species often uses the Linnaean system to reflect the changes. However, it becomes a difficult exercise to keep track of a new name change when still using what has become an outdated description (Tautz *et al.*, 2003).

It is not suggested that classification be DNA sequences replaces traditional taxonomy. A DNA sequence will remain the same for an individual regardless of its taxonomic affiliations. Tautz *et al*. (2002, 2003) proposed DNA taxonomy as a means of stabilising morphological taxonomy by grounding a species definition in an unchanging DNA sequence. Tautz *et al.*'s proposal has received global attention, with some expressions of support and some of challenge.

DNA taxonomy opponents have argued that limited funding would be directed towards it (Ronquist and Gardenfors, 2003) and that traditional taxonomy would be usurped by the more progressive science (Dunn, 2003;

Ebach and Holdrege, 2005; Lipscomb *et al.*, 2003). However, Gregory (2005) points out that DNA taxonomy and barcoding projects have been funded by agencies which traditionally do not support taxonomy. Moreover, international collaborative projects have multiple sources of funding which are more comparable with large-scale physics projects (Gregory, 2005). Morphological taxonomists acquire a unique and highly specialised skill set (Lee, 2000). However, their skills and expertise tend to be limited to routine identification and are under-utilised (Packer *et al.*, 2009; Tautz *et al.*, 2002). DNA taxonomy will require taxonomists to match existing information with DNA sequence (Gregory, 2005; Tautz *et al.*, 2003; Tautz *et al.*, 2002). Contrary to the opinions of some (e.g. Lipscomb *et al.*, 2003, Seberg *et al.*, 2003) research programs such as PEET[1] should help bridge the gap between the wide range of specimens awaiting description and the lack of trained taxonomists (Ronquist and Gardenfors, 2003).

Opponents also argue that introducing a mandatory inclusion of a DNA sequence into a species description (and hence generation from a new specimen) would be prohibitively expensive (Dunn, 2003; Seberg *et al.*, 2003). Yet, sequencing a target from a specimen is routine laboratory practice, barcodes can be generated in less than two hours (Ivanova *et al.*, 2009) with costs as little as $2[2] Canadian dollars (£1.28) per specimen. For DNA barcode projects, the bulk of costs are incurred from curation of specimens (Gregory, 2005).

Seberg *et al.* (2003) also note that specimens would be destroyed by DNA extraction procedures. For larger organisms, however, this is rarely an issue: only a minute portion needs to be sacrificed for DNA amplification (Tautz *et al.*, 2003). For insects, a leg will yield enough DNA for amplification (Hajibabaei *et al.*, 2005). Digital vouchering is a promising solution (De Ley and Bert, 2002; De Ley *et al.*, 2005; Tautz *et al.*, 2003) in those instances where whole specimens are needed to extract sufficient DNA (for example, nematodes and tardigrades).

---

[1] Partnerships for Enhancing Expertise in Taxonomy, USA, http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5451

[2] Barcode of Life Initiative (BOLI), 2008, http://www.barcoding.si.edu/PDF/CBOL-ABS%20Brochure%20-%20FINAL.pdf, accessed online May 2009

Monopolization of the information by developed nations (Seberg *et al.*, 2003) is another objection to centralizing DNA taxonomy. It is not made clear why this is considered an issue, however, as making projects and data available on the Internet would allow global access. As Minelli (2003) points out "Taxonomy, and science at large, will profit from an efficient way to access all nomenclatural information, ...". Tautz *et al*. (2003) and Gregory (2005) are quite clear that DNA taxonomy should be an evolution of traditional taxonomy.

As with traditional taxonomy, adapting existing practice to DNA taxonomy is likely to work well for larger animals. Descriptions of the majority of larger animals are fairly well accepted, although the advent of molecular information has yielded some significant surprises (Blaxter, 2003). Striking molecular differences, highlighting putative morphologically cryptic species, are also seen in beetles (Monaghan *et al.*, 2005), field voles (Hellborg *et al.*, 2005), scallops (Barucca *et al.*, 2004) and Neotropical frogs (Fouquet *et al.*, 2007).

The splitting of a single morphospecies into multiple molecular taxa by sequence data seems to be common among small eukaryotic species e.g. alga (Slapeta *et al.*, 2006), diatoms (Evans *et al.*, 2007), amphipods (Witt *et al.*, 2006), and *Chirononmus* (Diptera) (Pfenninger *et al.*, 2007). Nematode surveys repeatedly find high levels of cryptic diversity; organisms which morphologically appear identical but are molecularly divergent (Derycke *et al.*, 2005). For tardigrades, a molecular diversity survey suggested that the number of molecular taxa is much greater than indicated by the morphology of specimens (Blaxter *et al.*, 2004).

1.3 Operational Taxonomic Units vs. Species

Bacteriologists do not use a biological species concept because they cannot be applied to microorganisms without sex or sufficient diagnostic characters. Instead they refer to phylotypes as descriptors for different kinds of microbes (Sogin *et al.*, 2006). Taxa are identified by a short unique sequence from a target gene - a DNA barcode.

Using 16S rDNA sequences, environmental microbial diversity surveys have used numerous methods to define operational taxonomic units (OTUs) based

on sequence similarities (Sogin *et al.*, 2006) to infer taxon counts. Thus a difference of 3% between bacterial 16S rDNA sequences is widely regarded as indicative of different species (Quince *et al.*, 2008). Environmental surveys of microbes consistently reveal a few abundant taxa account for the majority of the diversity found, and there are also numerous highly diverse taxa of low abundance (Huber *et al.*, 2007; Sogin *et al.*, 2006). These results suggest that despite intensive sampling efforts, the diversity of the microbial world remains largely unknown (Huber *et al.*, 2007; Quince *et al.*, 2008; Sogin *et al.*, 2006)). This is situation is likely to be similar for meiofaunal taxa, as previous surveys have also found high levels of molecular diversity (Blaxter *et al.*, 2004; Floyd *et al.*, 2002).

Recognizing that a traditional species definition for meiofaunal taxa is not appropriate, for the above reasons, a DNA phylotype approach similar to that used by microbiologists could greatly increase the efficiency of the identification of meiofauna (Godfray, 2002). Morphological studies would remain an integral part of species description, and would be preceded by molecular screening of DNA. Sequences which do not correspond to previously defined OTUs could highlight taxa which may warrant further investigation. If required, paratype sequences from identified specimens may be incorporated to assign some level of taxonomic description to the OTUs. In most meiofaunal surveys, this would not be necessary, as OTUs can be used to describe diversity (Blaxter and Floyd, 2003).

1.4 Current Status of DNA Barcoding

There is a growing sense of urgency and responsibility to record life on Earth as humans influence the increasing rates of extinction. International programmes have sought to co-ordinate and drive efforts to complete a barcode catalogue. In the wake of the first wave of barcoding projects, the Consortium for the Barcode Of Life (CBOL, http://barcoding.si.edu) was launched in May 2004 with the aim of building a complete barcode library of all eukaryotic life. Through international collaborations of barcoding and taxonomic communities, CBOL has already initiated campaigns of All Birds Barcoding Initiative (ABBI, http://www.barcodingbirds.org), Fish Barcode Of Life Initiative (FISH-BOL, http://www.fishbol.org) and All Lepidopteran

Barcode Of Life Initiative (All-Leps, http://www.lepbarcoding.org) through international collaborations of barcoding and taxonomic communities. FISH-BOL and ABBI are global undertakings that aim to generate barcodes for 30,000 and 10,000 species respectively, by 2010 (Ratnasingham and Hebert, 2007). At the beginning of 2009, ABBI and FISH-BOL are 28% (2815 species) and 22% (6418 species) complete respectively. All-Leps has directed attention to lepidopterans of North America (USA and Canada with regional targeting of the Great Smokey Mountains National Park), Australia and Area de Conservación Guanacaste (Costa Rica), and the global families Geometridae, Saturniidae and Sphingidae. All-Leps currently has 12,259 species barcoded. With almost 180,000 described species and an estimated 300,000 undescribed species, there is still much to complete. Some campaigns are not limited to specific taxa. The Polar Barcode of Life Initiative (PolarBOL, http://www.polarbarcoding.org/) had collected over 20,000 specimens from terrestrial, marine and freshwater habitats. Sampling included nematodes, collembola, and marine taxa.

As the volume of data generated grows, the necessity for a centralized system of barcodes has been recognized. Originally designed as a workbench for a single barcoding facility, the Barcode of Life Data System (BOLD, http://www.barcodinglife.org) has since been adopted by the wider barcoding community (Ratnasingham and Hebert, 2007). BOLD provides a centralized system for the management and use of barcodes. To be formally accepted as a barcode sequence, specimens must fulfill several criteria such as species name, collection record (including dates and GPS locations), and sequence from the mitochondrial cytochrome oxidase c subunit I (COI) gene. To date, BOLD has just over 572,000 barcodes representing approximately 55,000 formally described species (accessed online April 2009). CBOL and BOLD represent some of the institutes providing support and organization for a worldwide barcode community. These are linked to other collaborators such as biodiversity organizations (e.g. GBIF[3], ToL[4]) and sequence data

[3] Global Biodiversity Information Facility, Denmark, http://www.gbif.org/
[4] Tree of Life, USA, http://tolweb.org/tree/

depositories (e.g. GenBank[5], EMBL[6], DDBJ[7]) and barcoding centres (e.g. Canadian Centre for DNA Barcoding, www.barcodeoflife.org).

1.5 Standardising DNA Barcodes

As with any scientific endeavour, standard protocols need to be developed and adopted to allow the possibility of integration among different studies. Whilst the extraction, amplification and sequencing of DNA are routine processes, a universal target (or targets) and analysis methods have yet to be agreed.

Some of the first meiofaunal barcode studies used small subunit ribosomal RNA (SSU or 18S) sequences (Floyd *et al.*, 2002). For example, work by Blaxter *et al.* (1998) showed that not only was it was easily recoverable from specimens, it also contained sufficient information, or signal, useful in the delineation of taxa. More recently, COI has been proposed as the standard barcoding target (Hebert *et al.*, 2003a; Hebert *et al.*, 2003b). Initially COI seemed to be an ideal target as there are conserved regions where apparently universal primers can bind, and divergent regions which contain signal (Hebert *et al.*, 2003a). Moreover, it is maternally inherited and, therefore, haploid within an organism. So sequencing is relatively straightforward as there is no heterzoygosity at the original locus. COI can identify birds (Hebert *et al.*, 2004b), butterflies (Hebert *et al.*, 2004a) and other organisms with a high degree of accuracy.

However, problems related to this particular gene have become apparent. Integration of mitochondrial DNA into the nuclear genome has been widely documented and can lead to amplification of non-functional fragments (nuclear mitochondrial pseudogenes or Numts) that yield incorrect phylogenetic and taxonomic relationships (Bensasson *et al.*, 2001; Tautz *et al.*, 2003). Whilst it may be suitable for delineating some species, COI barcode sequences lack the information required to construct accurate phylogenetic relationships (Hajibabaei *et al.*, 2006; Meyer and Paulay, 2005).

[5] US National Institutes of Health, http://www.ncbi.nlm.nih.gov/
[6] European Molecular Biology Laboratory, Germany, www.embl.org
[7] DNA Data Bank of Japan, Japan, http://www.ddbj.nig.ac.jp/

There are also technical issues with COI when compared with nuclear markers. It can be problematic to amplify consistently, and thus species-specific primers are often required (Lorenz *et al.*, 2005; Roe and Sperling, 2007). Whilst this may not be an issue for a study focusing on one or two taxa, for a broad diversity survey this could lead to profound systematic biases in taxon representations.

1.6 Standardised Analysis of DNA Barcodes

Delimitation of species through DNA barcodes is reliant on a barcoding gap, i.e. the qualitative difference between intraspecific and interspecific divergence (Figure 1.1).

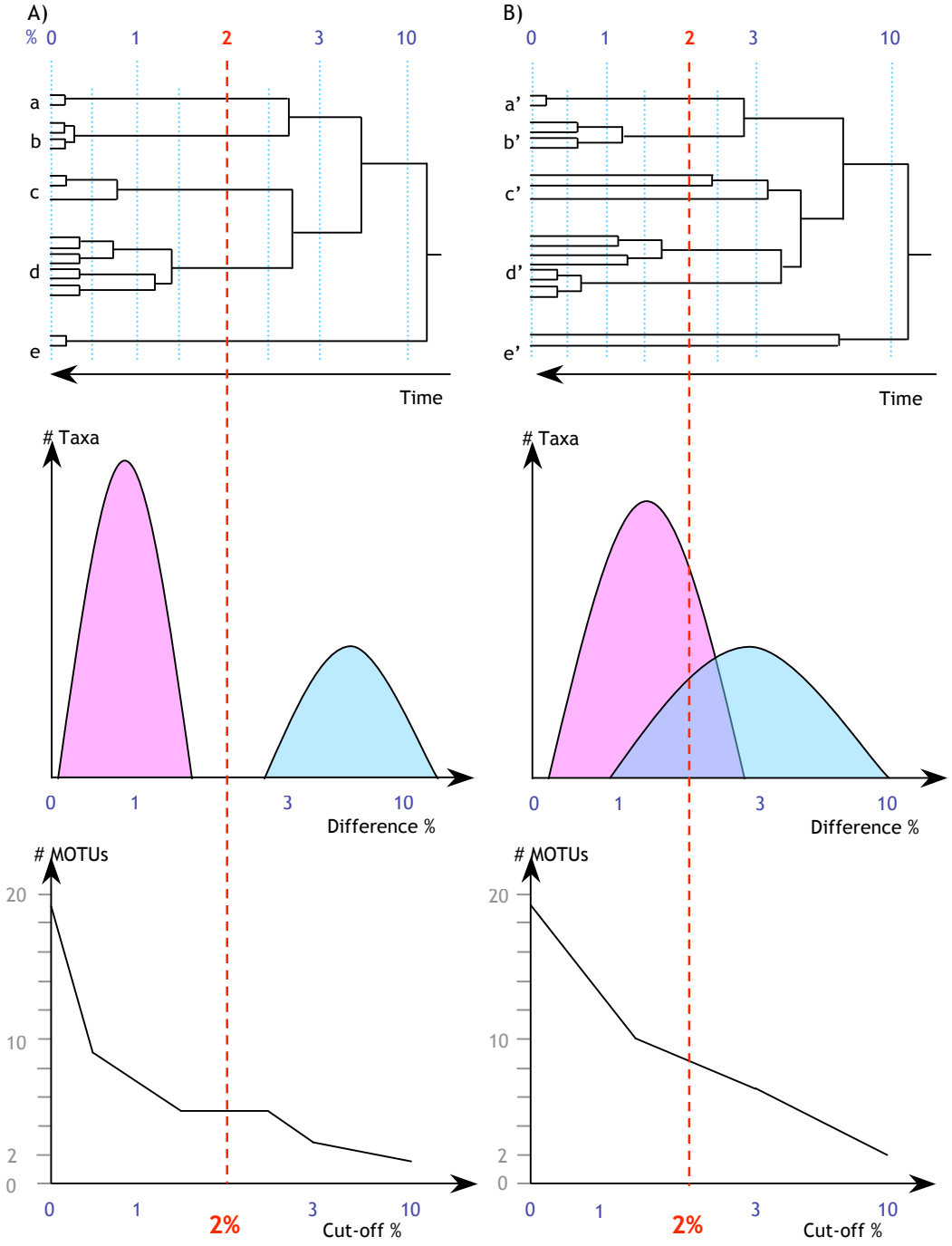
For any gene, individuals of the same species are expected to show a low level of difference, so reflecting a low level of intraspecific variation. As sister species represent different evolutionary lineages, over time gene sequences would be expected to diverge from their ancestral state. The amount of difference between the two species (interspecific variation) is thus an indication of the evolutionary distance between the two. For taxa which diverged millions of years ago, differences are expected to be much larger than intraspecific variation. Conversely, distinctions between intraspecific and interspecific variation would be less clear for species which diverged comparatively later.

Just as COI was put forward as the standard target, 2% sequence difference between COI sequences has been proposed as sufficient for species delimitation (Hebert *et al.*, 2003a; Hebert *et al.*, 2003b). In a survey of higher vertebrate taxa (fish, birds, mammals, amphibians and reptiles), sequences of mitochondrial cytochrome *b* (*cytb*) from the same species were shown to have bases which differed by less than 2%, whereas sequences from different species had differences greater than 2% (Johns and Avise, 1998). Johns and Avise also note that among the taxonomic ranks, there was little equivalency for genetic distances (1998); i.e. congeneric bird species had lower genetic distances than congeneric fish, mammal, amphibian and reptile species. This may well mean that it is not possible to establish a universal cut-off for all taxa.

However, for taxa not yet comprehensively sampled, the range of intraspecific and interspecific differences is largely unknown (Cognato, 2006). Even within some well studied invertebrate groups, the estimate of 2% difference is insufficient to delimit species (Cognato, 2006; Elias *et al.*, 2007; Wiemers and Fiedler, 2007). Populations within taxa with intraspecific differences which are higher than 2%, (e.g. salamanders at 7.8% (Vences *et al.*, 2005)) could be recognized as distinct species using this cut-off. Meiofauna, like some insect taxa, are likely to have short generation times. However, large effective populations are likely to have more genetic diversity and hence fewer fixed mutations. Different taxa will be affected more or less by these different factors, and this will limit the utility of a universal cut-off (Elias *et al.*, 2007).

As yet, there is no standard analysis method for DNA barcodes. The chosen method is often dependent on the purpose of the study, whether barcodes are being used to confirm identity to a species, or if they are being used to infer taxon diversity. Most confirmational investigations infer species identification by the placement of a specimen within a phylogenetic tree, e.g. butterflies (Hebert *et al.*, 2003a) and birds (Hebert *et al.*, 2004b). This applies a model of evolution to the data in order to infer phylogenetic relationships. Whilst this is relevant for investigating the deeper nodes among taxa, COI is known to contain limited phylogenetic information (Hajibabaei *et al.*, 2006; Meyer and Paulay, 2005). It has been suggested for a short barcode, it may be more appropriate to cluster sequences based on similarities in order for OTUs to be generated (Blaxter *et al.*, 2005; Floyd *et al.*, 2002; Sogin *et al.*, 2006).

Figure 1.1 Diagram to represent how a standard 2% difference relates to the "barcoding gap". Taxa (a-e, a'-e') are evolving in a clock like fashion but coalescents are at different times. **Part A**: Ideal situation, where all intraspecific comparisons (pink) have differences less than 2%. Differences of over 2% indicate interspecific differences (blue) and there is no overlap between the two ranges. This is reflected in a plateau phase in the MOTU number/ cut-off graph, where numbers of MOTUs are stable over a range of cut-off values around 2%. **Part B**: An alternative scenario, where different species have different levels of intra and interspecific variation such that the intra- and interspecific ranges overlap with no clear barcoding gap. There is no plateau phase as the number of MOTUs constantly decreases as the cut-off is increased. Adapted from Meyer and Paulay, 2005.



1.7 Aims of Thesis

This thesis seeks to address key issues associated with DNA barcoding of meiofauna.

Is COI a suitable target?

Whilst COI may be effective for identifying some species, its usefulness may be limited to larger animals which are well described. Meiofauna taxonomy generally lacks comprehensive taxon descriptions. Chapter 2 uses a thrips (Order Thysanoptera) dataset to assess whether COI barcodes can recover morphospecies. Chapter 3 assesses the utility of COI in comparison with two nuclear markers to define OTUs from a collection of nematodes, in the absence of any formal species identification.

Is there a universal cut-off for species delimitation?

In an attempt to standardize analysis of DNA barcodes, a sequence difference of 2% was proposed to be indicative of species boundaries. Again this was inferred from interspecific differences from larger animals, not meiofauna. Chapter 2 looks at a COI dataset for evidence of a barcoding gap. Chapter 3 investigates three different gene datasets and their potential barcoding gaps.

Is there an objective method to define taxa purely on sequence similarity?

Chapter 2 investigates results from two computer programs, challenged with the same dataset, to accurately recover the morphologically defined species. Chapter 4 compares OTU designation with BLAST results and putative morphological identification from unidentified specimens.

Can DNA taxonomy work as an identification tool for meiofauna?

If there are not enough trained taxonomists, non-specialists may be able to help with the detection of novel species. Chapter 4 develops protocols for the identification of meiofauna by integrating morphological detail with barcoding. Samples of meiofauna (nematodes, tardigrades, mites and copepods) were collected, preserved, and up to three barcodes were generated from each sample. Specimens were subjected to video capture and editing (VCE) (De Ley and Bert, 2002) before lysis. This enables morphological information to be reunited with sequence data and putative identifications made post-sequencing.

Chapter 2

Thrips COI Barcodes: Comparisons of the performance of two cluster analysis methods

2.1 Introduction

One of the problems encountered when trying to identify small organisms is their size, whereby key morphological features can be easily missed or misidentified. The vast majority of life on Earth has a body axis of less than 2 mm and many of the key morphological structures may be smaller than light microscopy resolution (De Ley and Bert, 2002). Even when a specimen is morphologically intact, it may be impossible to identify it to species in a quick, easy way. Invertebrate organisms are the most numerous animal taxa on Earth with many species poorly described. Those that have been recorded are likely to be a very small proportion of the true diversity (Blaxter, 2003; Wilson, 2003).

Thrips belong to the order Thysanoptera (phylum Arthropoda, class Insecta) and are small winged insects ranging from 0.5 mm to 15 mm in length (Gullan and Cranston, 2005). Worldwide, there are over 5000 recorded species of thrips (Crespi *et al.*, 1996; Brunner *et al.*, 2004; Inoue and Sakurai, 2007). Traditionally adult morphological characters and host plants have been used to classify two suborders; Terebrantia consisting of eight families, and Tubulifera with a single family (Brunner *et al.*, 2004; Crespi *et al.*, 1996). Although the monophyly of Thysanoptera is supported by morphological and molecular evidence (Brunner *et al.*, 2002; Inoue and Sakurai, 2007), the relationships among the nine families in the sub-orders are unresolved using morphological traits (Mound *et al.*, 1980).

Around 100 species have been identified as pest-species (Brunner *et al.*, 2002) according to feeding behaviour, causing damage to plants, or as disease vectors (Toda and Murai, 2007). Easily over-looked, these small organisms have been transported across the world following trade routes for vegetables and ornamental flowers (Brunner *et al.*, 2002). Removing a species from its native habitat into a novel environment can lead to the species becoming invasive and damaging its new local environment (Brunner *et al.*, 2002; Toda

and Murai, 2007). Although thrips can damage crops, different species are of varying importance to agriculture. In particular, *Thrips tabaci* Lindeman is a polyphagous species as well as being a vector of tomato spotted wilt virus (TSWV) and is therefore of economic importance (Brunner *et al.*, 2004; Toda and Murai, 2007). Different species vary in their importance to us, particularly pest species that are subject to quarantine regulations (Armstrong and Ball, 2005) and therefore need to be correctly identified. Traditionally, thrips identification uses keys mainly based on adult morphology from type specimens (Mound *et al.*, 1980). Identifying specimens this way relies on experts and takes time (Tautz *et al.*, 2003). If a specimen is a larva or damaged, important morphological structures used for identification are likely to be absent. When sampling, it is highly unlikely that all specimens will be adults (Hosseini *et al.*, 2007), making identification keys insufficient at identifying all specimens sampled.
A quick and reliable method to identify specimens, regardless of condition, size or life stage, without the need for taxonomic experts, would facilitate routine identification and quarantining of pest thrips species.

Molecular Barcoding

The introduction of molecular diagnostic tools has aided the accuracy and speed of species identification. Molecular barcoding, using PCR to obtain short DNA sequences to identify specimens, has two applications. Firstly, it can be used to identify an unknown specimen by comparison of a short DNA sequence to a comprehensive data set of sequences from identified species (Moritz and Cicero, 2004). Alternatively, DNA barcoding can be used to aid species discovery (Meyer and Paulay, 2005). Initially molecular barcodes were used to identify particular species (Gasser *et al.*, 1994) and methods used were often restricted to the study. More recently, however, in the widening gap between diagnostic needs and trained taxonomists (Armstrong and Ball, 2005; Tautz *et al.*, 2003) molecular barcoding for species confirmation and species discovery has become widespread, and the need for a universal approach has long been recognised (Moritz and Cicero, 2004). Intraspecific (within species) variation in DNA sequences is expected to be small. Differences between species (interspecific variation) should be greater

and relate to the length of time of divergence between species. Although the magnitude of both intra- and interspecific variation will vary depending on the study taxon, in a perfect barcoding world there should be no overlap between the two (Meyer and Paulay, 2005). This gap is referred to as the 'barcoding gap' and separates the coalescent of individual variation from the birth-death process of species' intraspecific divergences.

Of course there are several situations where a barcoding gap may not exist. For example, where two distant populations of a species are genetically distinct due to limited gene flow, barcoding would incorrectly indicate a gap (Wiemers and Fiedler, 2007). In addition to this false positive problem, false negatives may be found in barcoding where no barcoding gap is found, e.g. when there is little sequence variation found in the barcoding gene. This could be true for very closely related species where ancestral polymorphism is still segregating or hybridisation is maintaining identity in both species (Trewick, 2008).

Initial work by Johns and Avise (1998) on mitochondrial cytochrome *b* demonstrated that different vertebrate classes showed different levels of variation when genera within a class were compared. Amphibians, reptiles and fish showed large distances when compared to mammals and birds (which have the smallest distances) (Johns and Avise, 1998). The study suggested also that a mean difference of more than 2% between sequences would be sufficient to distinguish between vertebrate species.

The mitochondrial gene cytochrome oxidase *c* subunit I (COI) has been proposed as the standard for molecular barcoding (Hebert *et al.*, 2003a). Early studies suggested that it could consistently and faithfully recover species based on differences in sequences (Hebert *et al.*, 2003b). Recent studies found a similar pattern to Johns and Avise (1998). Hebert et al. (2004a) used 437 COI sequences representing 260 bird species and found the average intraspecific difference was 0.43%. However, Vences *et al.* (2005) found intraspecific variation for mantillid frogs and salamanders was as high as 10-14% and 7.8% respectively. At this level, intraspecific variation overlaps with interspecific variation so that species delineation was difficult. These studies all investigated well-defined and described vertebrate species. Invertebrates, on the other hand, are numerous, often have large effective population sizes

and high speciation rates (Elias *et al.*, 2007). These two factors are likely to inhibit the usefulness of barcodes to identify invertebrate species with either too much or too little variation respectively.

Initial work on invertebrate barcodes intimated that 2% difference in COI sequences would separate species. Work by Hebert *et al*. (2003a) on COI analysis of eight insect orders and 200 lepidopteran species suggested that 2% was capable of delimiting species. However, later work on other butterfly species (Elias *et al.*, 2007; Wiemers and Fiedler, 2007) failed to show such confidence in the ability of barcodes to identify species. Moreover, a recent investigation of New Zealand grasshoppers (order Orthoptera) failed to find any barcoding gap or matches between molecular and accepted morphological taxonomy (Trewick, 2008). Identification of Dipterans using DNA barcodes was also problematic (Meier *et al.*, 2006).

The main issue of earlier studies has been the insufficient sampling of taxa (Trewick, 2008). If only one or two individuals were sampled within a species (Hebert *et al.*, 2003a), then it would not be possible to estimate the range of intraspecific variation and the likelihood of creating a false barcoding gap increased (Meyer and Paulay, 2005; Trewick, 2008).

As well as the issues discussed above, there are other potential drawbacks with COI. Primarily, it is maternally inherited and so can only ever reflect maternal evolution (Rubinoff, 2006). There is also evidence to suggest inherited symbionts can affect the variation of the mitochondrial genome within a species (Hurst and Jiggins, 2005). Moreover, there are technical issues relating to the 'universal' nature of primers for the mitochondrion gene target. Where variability is high, multiple primers are required to recover COI targets from all specimens (contradicting results from Hebert *et al.*, 2003a) and taxon specific primers must be designed. COI is also more difficult to amplify from some specimens and has a lower recovery rate when compared with other genes, such as the nuclear ribosomal large subunit (LSU or 28S) (see Chapter 3 of this thesis).

Turning sequences into taxa

If COI was to be adopted as the standard for barcoding, there should also be a standard method for derivation of taxa from barcodes. Sequences may be compared by a simple BLAST search (Altschul *et al.*, 1997) using the similarity score to define taxa. Distances between sequences may also be used to construct phylogenetic trees to delineate taxa, using branch length as a measure of relatedness. This method lacks objective criteria to designate taxa as clades are defined by eye.

Pons *et al*. (2006), delineated specimens of tiger beetles by identifying putative species based on branch lengths of clades, from a mtDNA phylogenetic tree. Using a molecular clock to 'best-fit' the data, a change in the rate of branching was assumed to be indicative of a species boundary. This method does not require species or populations to be defined a priori but it was assumed that each geographic sampling site represented a separate population unless morphological differences were observed and that the sampling regime was a good reflection of the total diversity (Pons *et al.*, 2006). However, only using mtDNA (in effect, a single locus) would fail to distinguish recently diverged lineages (Elias *et al.*, 2007 2008) or recently derived geographic populations (Pons *et al.*, 2006). If this method was to be used as a way to infer species from barcodes (Pons *et al.*, 2006) then taxa need to be extensively sampled. Pons *et al.* concede that "the extent of population sampling...may rarely be complete" (2006). Therefore, this method is unsuitable when trying to cluster high throughput barcode data that are likely to contain sequences from incompletely sampled species, isolated clades or populations that share gene flow (Lohse, 2009).

MOTU_define.pl

MOTU_define.pl (M. Blaxter, J. Mann and R. Floyd, unpublished, see http://www.nematode/org/bioinformatics/ for download) has been developed to cluster sequences into molecular operational taxonomic units (MOTUs) independent of phylogenetics. During MOTU_define.pl analysis, a sequence is compared pair-wise to all other sequences in the data set in a random order (the primary run). The program generates a local database of

previously defined MOTU. MOTU_define.pl will ask if a query sequence matches any other sequences in the database with less than *x* b difference (the cut-off value) over more than a minimum overlap of *y* bases (Blaxter *et al.*, 2005; Appendix 2.1). If the sequence matches a previously defined MOTU by less than the cut-off and along the overlap, then it is assigned to that MOTU. If it does not match, it forms a new MOTU and is given the next sequential MOTU identifier. A new sequence is then picked at random, and the process repeated. The user sets the cut-off and minimum overlap. Membership of a MOTU as defined by MOTU_define.pl, can be affected by the order in which the sequences are added (Blaxter *et al.*, 2005). Therefore, re-sampling is important to investigate the variability of MOTU classification at any cut-off.

MOTU_define.pl does not establish the relatedness of MOTUs (although this can be investigated by observing changes in membership over different cut-offs). It can deal with isolated species, populations sharing gene flow and incomplete sampling of populations. Moreover, it is incremental, so that new data can be added to previously defined MOTUs without the need to start analysis from the beginning.

DOTUR

Defining Operational Taxonomic Units and Richness (DOTUR) is another method for assigning sequences to defined operational taxonomic units (here called DOTUs) (Schloss and Handelsman, 2005). DOTUR defines taxa by clustering sequences based on distances derived from an alignment of the sequences. DOTUR uses three methods for clustering sequences, furthest, nearest and average neighbour. Furthest neighbour clustering only adds a sequence to a DOTU if it is sufficiently similar to all other member sequences in the DOTU, otherwise it will seed a new DOTU. Nearest neighbour adds a sequence to a DOTU if it is similar to any sequence in it. This means that if a DOTU has many sequences, the difference between the two most distant sequences within it could be quite large. Nearest neighbour clustering would be expected to define a similar number of DOTUs to MOTUs as both use a similar method to assign sequences to operational taxonomic units (OTU). Average neighbour clustering is an intermediate between the two methods.

As well as different clustering methods, it is also possible to construct DNA distance matrices using different models of evolution. The Jukes-Cantor model assumes that there is no difference between transition and transversion rates when comparing sequences. COI is a protein-coding gene and thus there are likely to be differences between transversional and transitional rates, as modelled by the Kimura "2-Parameter" model.

What cut-off should be used to define taxa?

A difference of 2% between sequences has been proposed as the cut-off for defining species (Hebert *et al.*, 2003a). However this universal cut-off is not suitable for all animals as inter- and intraspecific distances vary among genera (DeSalle *et al.*, 2005; Vences *et al.*, 2005). Rather than assuming 2% is sufficient as a cut-off value, both MOTU_define.pl and DOTUR allow the user to investigate the clustering of sequences over multiple cut-offs, and so reveal a barcoding gap if one is present. Ideally, multiple lines of evidence, such as multiple genes with different modes of inheritance (Elias *et al.*, 2007), and more than one analysis method should be used to support the morphological and sequence-based definition of species.

In this investigation, a large dataset of partial COI sequences was used to assess the ability of MOTU_define.pl and DOTUR to define thrips taxa based on sequence data. The dataset consisted of 332 specimens that had been morphologically identified and then sequenced for the 5' region of COI. The sequence data were analysed, independent of morphological designation, to define OTUs. The OTUs were then compared with morphological species designation (morpho-species) to test the ability of barcode OTU methods to recover morphologically identified taxa. A phylogenetic analysis was also performed as an independent check of OTUs defined.

2.2 Methods

Dr. Dom Collins of Central Science Laboratories (CSL), York, collected and identified 332 specimens of thrips representing 46 species (order Thysanoptera, sub-orders Terebrantia and Tubulifera) (Table 2.1). Rachel Glover at CSL performed PCR, sequencing and trace analysis and trimming. Genomic DNA was extracted from whole specimens, ranging from first instar larvae to adult specimens. The mitochondrial COI gene was amplified using the PCR primer pairs MTD 7.2F (5'-ATT AGG AGC HCC HGA YAT AGC ATT-3') and MTD 9.2R (5'-CAG GCA AGA TTA AAA TAT AAA CTT CTG-3') or LCO1490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') and C1-N-2191 (5'-CCC GGT AAA ATT AAA ATA TAA ACT TC-3') to amplify approximately 500 b and 700 b from the 5' region of the COI gene respectively. Positive results were cleaned and sequenced in both directions using standard protocols. Bases were called and edited by eye by Rachel Glover.

Sequences ranged from 353 to 461 base pairs (bp) in length with a mean of 454.7 bp, and twenty sequences were fewer than 400 bp. Of the 46 morphological species, three (*Haplothrips leucanthemi*, *Merothrips floridensis* and *Odontothrips ignobilis*) had only one specimen and one sequence. The rest of the species had multiple specimens, with a maximum of 37 specimens for *Thrips palmi*.

Sequences were received as a catenated fasta file, and were processed using MOTU_define.pl and DOTUR and a phylogenetic analysis was performed with the addition of a non-thrips sequence as an outgroup. Sixty nine of the sequences have been deposited in GenBank by CSL (Table 2.1).

Table 2.1 Taxonomic summary of 332 thrips (Insecta, Thysanoptera) sequences provided by CSL with CSL ID numbers listed. ID numbers in italics indicate the sequence has been deposited in GenBank and the accession numbers are listed. Numbers of sequences from each sub-order are displayed.

Sub-order	Family	Sub-family	Genus	Species	*n*	CSL ID	GenBank accession number
Terebrantia 307	Merothripidae		*Merothrips*	*floridensis*	1	T231	
	Aeolothripidae		*Aeolothrips*	*albicinctus*	3	T6, T139, T147	
	Thripidae	Panchaetothripinae	*Hercinothrips*	*femoralis*	3	T239-T241	
			Parthenothrips	*dracaenae*	3	T232-T234	
		Thripinae	*Ceratothrips*	*ericae*	3	T334, T335, T336	
			Chirothrips	*manicatus*	10	T80, T81-T83, T88-T90, T337-T339	
				meridionalis	5	T69-T73	
			Dendrothrips	*degeeri*	4	T8, T18, T137, T140	
			Echinothrips	*americanus*	6	T225-T230	
			Frankliniella	*intonsa*	8	T16, T24, T215-T217, T311-T313	
				occidentalis	26	*T32, T96, T128-T130, T164-T166, T181, T185, T186, T193* T106, T108, T118, T125, T163, T184, T194, T210, T265-T267, T287-T289	AM931992, AM932005, AM932016-AM932018, AM932021-AM932023, AM932025-AM932027, AM932029
				schultzei	5	T127, T167-T170	
			Hydatothrips	*adolfifriderici*	5	T126, T156-T159	
			Kakothrips	*robustus*	3	T323-T325	
			Limothrips	*denticornis*	2	T359, T360	
			Odontothrips	*biuncus*	2	T384, T385	
				ignobilis	1	T351	
				loti	2	T380, T381	
				phalaratus	4	T47-T50	
				ulicis	9	T1, T3, T44-T46, T132, T134,	

			T143, T150	
Oxythrips	*ajugae*	3	T5, T145, T151	
Scirtothrips	*dorsalis*	9	T247-T249, T262-T264, T361-T363	
Sericothrips	*staphylinus*	7	T40-T43, T346-T348	
Thrips	*angusticeps*	3	T299-T301	
	brevicornis	3	T314-T316	
	flavus	29	*T12, T14, T15, T28-T30, T105, T179, T192, T203, T207-T209* T22, T107, T110, T111, T120, T178, T187-T191, T195, T202, T340-T342	AM931983-AM931985, AM931988-AM931990, AM932009, AM932024, AM932028, AM932034, AM932038-AM932040
	fuscipennis	3	T308-T310	
	hawaiiensis	7	T364-T370	
	major	11	T7, T19, T23, T138, T198, T280-T282, T305-T307	
	minutissimus	6	T2, T4, T144, T146, T148, T149	
	nigropilosus	13	*T26, T27, T31, T37, T38, T235, T236, T271-T273, T277-T279*	AM931986, AM931987, AM931991, AM931996, AM931997, AM932044, AM932045, AM932047-AM932052
	palmi	37	*T33-T35, T58-T60, T84-T86, T115, T117, T196, T197, T199, T201, T204-T206, T211, T214, T238* T36, T87, T112, T114, T116, T200, T212, T213, T237, T268-T270, T274-T276	AM931993-AM931995, AM931999-AM932004, AM932010, AM932011, AM932030-AM932033, AM932035-AM932037, AM932041, AM932042, AM932046
	sambuci	2	T56, T57	
	tabaci	34	*T39, T97, T99, T100, T121-T124, T160, T162, T223* T25, T98, T113, T119, T161, T171-T173, T180, T182, T183,	AM931998, AM932006-AM932008, AM932012-AM932015, AM932019, AM932020, AM932043

					T222, T224, T283-T285, T296-T298, T321, T322, T352-T354
			trehernei	8	T11, T91-T95, T136, T142
			urticae	8	T101-T104, T317-T320
			validus	7	T9, T20, T135, T141, T355-T357
			vulgatissimus	12	T13, T21, T61-T64, T328-T330, T343-T345
Tubulifera 25	Phlaeothripidae	*Cephalothrips*	*monilicornis*	4	T242, T243, T326, T327
		Haplothrips	*aculeatus*	4	T51-T54
			cenchricola	3	T244-T246
			distinguendus	4	T286, T290-T292
			leucanthemi	1	T10
			setiger	3	T394-T396
			statices	3	T331-T333
			subtilissimus	3	T302-T304
				332	

2.2.1 MOTU_define.pl

MOTU_define.pl version 2.08 (M. Blaxter, J. Mann and R. Floyd, unpublished) was used to investigate the assignment of thrips sequences to molecular operational taxonomic units (MOTUs).
For this analysis, a range of cut-offs was used (0-15, 20, 25, 30, 35, 40, 45, 50 bp), equivalent to 0 – 10.9% divergence, and 100 re-samplings at each cut-off were performed. The shortest sequence was T308_Thrips_fuscipennis (353 bp), so minimum sequence length was set to 350 bp, with a minimum overlap of 210 bp. For each cut-off the mean (and standard deviation) of the number of MOTUs defined was calculated from the re-samples.
The congruence between MOTU and morpho-species was assessed by examining the monospecificity of MOTUs (MOTUs consisting of sequences from a single morphological species) over the cut-off range. Complete monospecificity would be attained when all morpho-species sequences were found in a single MOTU with the exclusion of other morpho-species. At low cut-off values, individual variation would result in sequences from the same morpho-species being clustered into multiple MOTUs, so many MOTUs will be monospecific, but the number of complete monospecific MOTUs should be low. As the cut-off increases, morpho-species sequences should coalesce, and the number of complete monospecific MOTUs would be expected to rise. This proportion should continue to rise until an optimum cut-off is reached and a maximum number of complete monospecific MOTUs found. After this peak, the high cut-off value will produce MOTUs that represent groups of species (genera), so the proportion of monospecific MOTUs (complete or otherwise) should start to fall. Investigating the monospecificity of MOTUs will highlight the accuracy of morphological identification relating to molecular differences in taxa.

2.2.2 DOTUR

Using ClustalX, at default settings, an alignment was constructed from the 332 thrips COI sequences from CSL. Two DNA distance matrices were produced using Jukes-Cantor and Kimura “2-Parameter” models to estimate

distances. For each distance matrix, DOTUR was performed using all three clustering methods as each method was likely to define different numbers of DOTUs. DOTUR reports the distance at which the number of DOTUs changes. The detail at which the distance was reported (the precision level "-p") was set to 1000 so that distance was reported to 3 decimal places as a percentage of the mean sequence length. The number of DOTUs found at different cut-offs was compared with MOTU_define.pl results. The data were also interrogated for evidence for a barcoding gap and congruence with morphological species.

2.2.3 Phylogenetic analysis

Previous analysis of thrips sequences have tended to concentrate on specific taxa and have used another thrips sequence as an outgroup and (Crespi *et al.*, 1998; Inoue and Sakurai, 2007; McLeish *et al.*, 2006; Mound and Morris, 2001). The Tree Of Life (http://tolweb.org), places thrips in the Hemipteroid assemblage, and recent work on SSU data for the assemblage suggests that thrips are sister taxon to Hemiptera (Murrell and Barker, 2005; Yoshizawa and Johnson, 2005). A COI sequence from *Acyrthosiphon pisum* (order Hemiptera), GenBank accession number EU071328, was added as an outgroup to the ClustalX alignment (Murrell and Barker, 2005). MrModeltest2 version 2.3 (Nylander., J. A. A; see http://www.abc.se/~nylander/ for download) was used to establish that a GTR+G model best fitted the data using hierarchical Likelihood Ratio tests. This was used to generate a maximum likelihood (ML) tree in PAUP 4.0b10 (Swofford, 2002). Two thousand bootstrap re-samples were preformed to test support for branches the ML tree.

2.3 Results

2.3.1 MOTU_define.pl results

The primary run for each cut-off was used to designate the name of a MOTU (e.g. MOTU0002). The mean number of MOTU defined at 0% (0 bp) cut-off was 197.74 ±1.28 (Figure 2.1). The number of MOTU decreased to 146.99 ±2.21 at 0.2% (1 bp) then 117.72 ±2.23 at 0.4% (2 bp). Increasing the cut-off from 0 - 1.98% of mean sequence length (0 - 9 bp) resulted in a sharp fall in the number of MOTUs defined. From 2.2 - 5.5% (10 - 25 bp) cut-offs, the mean number of MOTUs decreased only slightly from 54.21 ±1.48 to 43.3 ±0.69 MOTUs. Identifying a plateau in MOTU number, where there is little or no change in the number defined, is indicative of a barcoding gap. The smallest decrease in the number of MOTU defined was between 4.4 - 5.5% cut-offs (20 - 25 bp) where the number of MOTUs drops from 44.16 ±0.44 to 43.3 ±0.69. A further drop in the number of MOTU defined was seen as the cut-off was increased to 11% (Figure 2.1).

As 2% has been proposed as sufficient for designating arthropod taxa when working with barcodes (Hebert *et al.*, 2003a), this was the level at which the membership of MOTUs was investigated for the thrips data. Two percent of the mean sequence length (454.7 bp) was 9.09 bp; the nearest cut-off was 9 bp (1.98%). At this level, the mean number of MOTUs was 58.99 ±1.57. The primary run reported 59 MOTUs.

I examined the stability of these MOTU to the random-addition order re-sampling process. Some MOTU were robust and were consistently equivalent to the primary run (shaded cells, Table 2.2). Sequences such as T231 form a singleton MOTU in all of the 100 re-samples (MOTU0045, Table 2.2). Other MOTUs were not robust because the majority of the re-samples failed to produce MOTUs equivalent to the primary run.

Figure 2.1 Mean number of MOTUs generated using MOTU_define.pl, cut-offs are given as a percentage of the mean sequence length. Means and standard deviation calculated from 100 re-samples.

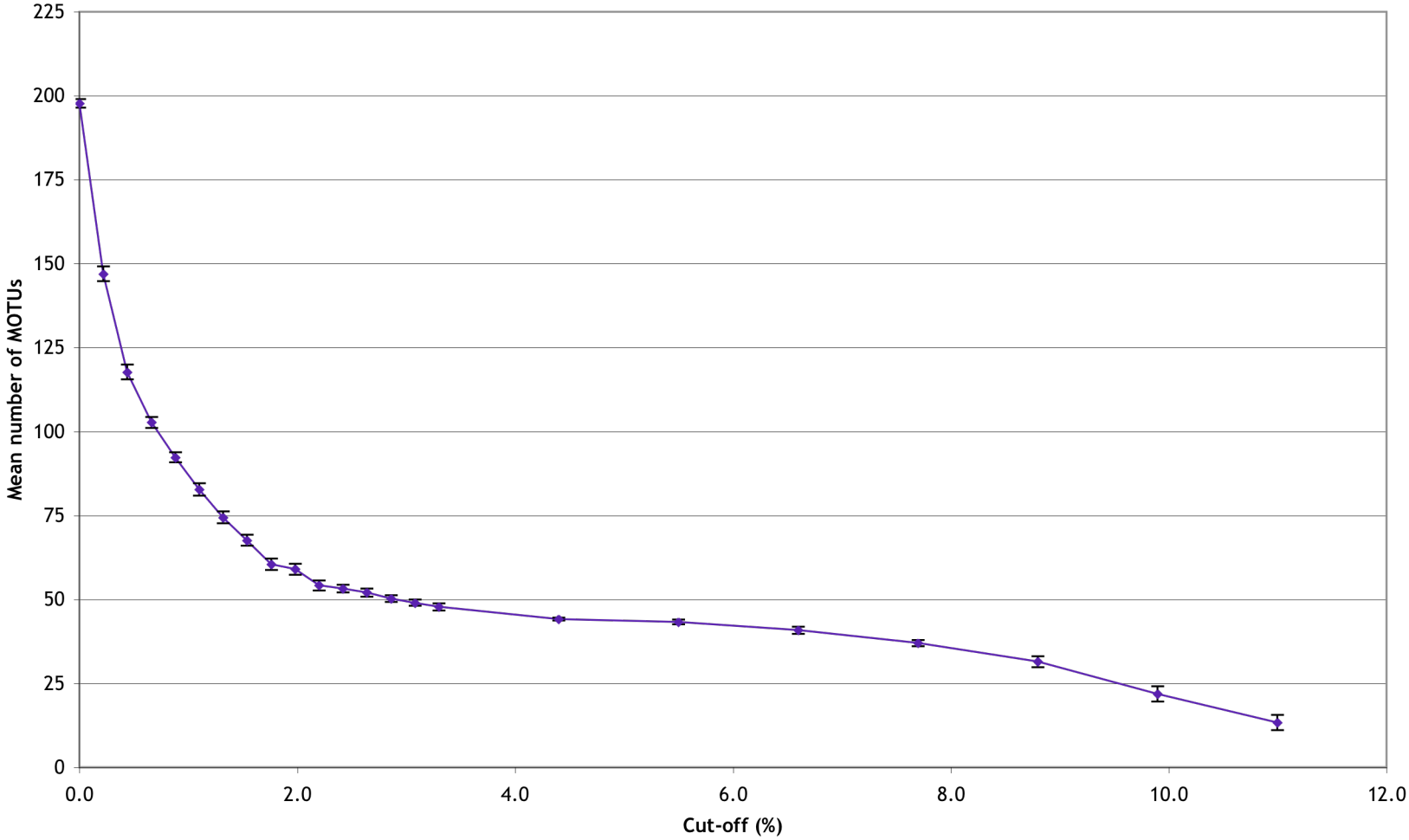


Table 2.2 MOTU designations of sequences at 9 b for the primary run compared with the MOTU designation from 100 re-samples. MOTU ID and member sequences were defined from the primary run with the number of sequences in each MOTU (*n*) and the morphological species listed. MOTUs from the re-samples were classed as **E = Equivalent** (if they had exactly the same members as the primary MOTU); **S = Split** (if the primary MOTU had been simply split to produce multiple MOTUs, e.g. one MOTU containing sequences *a*, *b* and *c* could split to form two MOTUs (*a*+*b*, *c*) or three MOTUs (*a*, *b*, *c*)); **J = Join** (if the MOTU had more sequences than in the primary run): **C = Complex** (if the resample MOTU had been split and joined by other sequences when compared to the primary run). * Indicates the join was a result of the MOTU merging with any of the sequences from MOTU0016 (T126, T156, T157, T158 or T159). MOTUs that remained unchanged in all of the re-samples are shaded.

9 b Primary Run				**Re-samples**			
MOTU_ID	**Member sequences**	**n**	**Species**	**E**	**S**	**J**	**C**
MOTU59	T351	1	*Od. ignobilis*	100	0	0	0
MOTU58	T395 T394 T396	3	*Ha. setiger*	96	0	4*	0
MOTU57	T307	1	*T. major*	62	0	38	0
MOTU56	T314	1	*T. brevicornis*	100	0	0	0
MOTU55	T63	1	*T. vulgatissimus*	100	0	0	0
MOTU54	T316 T315	2	*T. brevicornis*	98	0	2*	0
MOTU53	T130 T128	2	*F. occidentalis*	97	0	3*	0
MOTU52	T6 T147 T139	3	*A. albicinctus*	98	0	2*	0
MOTU51	T145	1	*Ox. ajugae*	98	0	2*	0
MOTU50	T225 T228 T226 T227 T230 T229	6	*E. americanus*	94	0	6*	0
MOTU49	T323 T324 T325	3	*K. robustus*	97	0	3*	0
MOTU48	T52 T54 T51 T53	4	*Ha. aculeatus*	97	0	3*	0
MOTU47	T10 T331 T333 T332	4	*Ha. leucanthemi; Ha. statices*	96	0	4*	0
MOTU46	T43 T348 T42 T346 T347 T41	6	*Se. staphylinus*	6	0	94	0
MOTU45	T231	1	*M. floridensis*	100	0	0	0
MOTU44	T360	1	*L. denticornis*	99	0	1*	0
MOTU43	T3 T134 T1 T143 T45 T46 T150 T132 T44	9	*Od. ulicis*	95	0	5*	0
MOTU42	T22	1	*T. flavus*	98	0	2*	0
MOTU41	T73 T71 T70 T69 T72	5	*Ch. meridionalis*	95	0	5*	0
MOTU40	T238 T275 T270 T237 T276 T60	6	*T. palmi*	22	0	74, 4*	0
MOTU39	T291 T286 T290 T292	4	*Ha. distinguendus*	96	0	4*	0
MOTU38	T335 T334 T336	3	*Cer. ericae*	98	0	2*	0
MOTU37	T380 T381	2	*Od. loti*	99	0	1*	0
MOTU36	T49 T47 T48 T50	4	*Od. phalaratus*	94	0	4*	0
MOTU35	T283 T160 T284 T183 T285 T180 T171 T173 T172 T352 T353 T182 T354	13	*T. tabaci*	48	40	8*	4
MOTU34	T359	1	*L. denticornis*	100	0	0	0
MOTU33	T299 T301 T300	3	*T. angusticeps*	95	0	5*	0

MOTU32	T81 T90 T88 T83 T80	5	*Ch. manicatus*	92	0	8*	0
MOTU31	T364 T366 T367 T368 T365 T369 T370	7	*T. hawaiiensis*	88	6	6*	0
MOTU30	T103 T319 T317 T104 T101 T102 T320 T318	8	*T. urticae*	93	0	7*	0
MOTU29	T311 T313 T215 T217 T24 T216 T312 T16	8	*F. intonsa*	93	0	7*	0
MOTU28	T385 T384	2	*Od. biuncus*	97	0	3*	0
MOTU27	T19 T306 T198 T7 T138 T282 T23 T280 T281 T305	10	*T. major*	53	0	38, 9*	0
MOTU26	T244 T246 T245	3	*Ha. cenchricola*	98	0	2*	0
MOTU25	T232 T233 T234	3	*P. dracaenae*	97	0	3*	0
MOTU24	T168 T170 T169 T127 T167	5	*F. schultzei*	94	0	6*	0
MOTU23	T40	1	*Se. staphylinus*	6	0	94	0
MOTU22	T18 T8 T140 T137	4	*D. degeeri*	95	0	5*	0
MOTU21	T5 T151	2	*Ox. ajugae*	96	0	4*	0
MOTU20	T309 T57 T310 T308 T56	5	*T. fuscipennis; T. sambuci*	93	0	7*	0
MOTU19	T239 T240 T241	3	*He. femoralis*	99	0	1*	0
MOTU18	T304 T302 T303	3	*Ha. subtilissimus*	97	0	3*	0
MOTU17	T330 T343 T21 T345 T344 T329 T64 T13 T328 T61 T62	11	*T. vulgatissimus*	91	0	9*	0
MOTU16	T157 T126 T158 T156 T159	5	*Hy. adolfifriderici*	9	0	0	91
MOTU15	T248 T263 T262 T247 T363 T362 T264 T249	8	*Sc. dorsalis*	6	0	81, 13*	0
MOTU14	T32	1	*F. occidentalis*	100	0	0	0
MOTU13	T136 T93 T92 T94 T11 T91 T142 T95	8	*T. trehernei; T. trehernei-pelikani*	92	0	8*	0
MOTU12	T268	1	*T. palmi*	25	0	74, 1*	0
MOTU11	T148 T146 T149 T4 T144 T2	6	*T. minutissimus*	95	0	5*	0
MOTU10	T84 T212 T87 T117 T36 T269 T206 T122 T85 T196 T34 T213 T199 T214 T211 T35 T86 T201 T114 T204 T274 T112 T116 T58 T197 T205 T115 T200 T59 T33	30	*T. palmi*	58	17	21*	4
MOTU09	T326 T243 T242 T327	4	*Cep. monilicornis*	93	0	7*	0
MOTU08	T265 T129 T287 T210 T185 T96 T108 T125 T289 T166 T184 T193 T181 T165 T163 T288 T118 T266 T186 T267 T194 T106 T164	23	*F. occidentalis*	19	53	3*	25
MOTU07	T97 T124 T39 T121 T119 T322 T321 T25 T224 T99 T297 T162 T161 T223 T123 T113 T222 T298 T100 T98 T296	21	*T. tabaci*	72	0	28*	0

MOTU06	T356 T9 T141 T135 T20 T355 T357	7	*T. validus*	88	0	12*	0
MOTU05	T82 T338 T337 T339 T89	5	*Ch. manicatus*	91	0	9*	0
MOTU04	T361	1	*Sc. dorsalis*	8	0	92	0
MOTU03	T189 T105 T190 T12 T209 T192 T342 T195 T107 T110 T341 T120 T188 T208 T14 T111 T191 T202 T187 T340 T30 T28 T179 T178 T203 T29	26	*T. flavus*	75	4	18*	3
MOTU02	T235 T236 T27 T277 T271 T272 T279 T273 T278 T26 T31 T37 T38	13	*T. nigropilosus*	20	67	5*	8
MOTU01	T207 T15	2	*T. flavus*	100	0	0	0

Investigation of the membership of MOTUs defined using 9 bp (2%) highlighted the effect of addition order and therefore the usefulness of re-sampling. The primary run defined MOTU0016 containing sequences T126, T156, T157, T158 and T159 (Table 2.2). However, examination of the re-samples showed there were only nine re-sampling events where the same MOTU was recovered. From the re-sample results, it was clear that MOTU0016 was unstable at this cut-off value. The re-sample data demonstrated that four of the sequences distributed among other MOTUs and a singleton MOTU (containing either T157 or T158), was more stable. If the data set consists of closely related species, and the order in which they were added influences the results, the re-sampling function represents a valuable tool for investigating the average membership of a sequence to any MOTU (Table 2.2).

Other examples highlighted where the primary run did not reflect the commonest outcome compared with the re-sample data. MOTU0046 was only recovered six times in the re-samples. A MOTU merged from MOTU0046 and MOTU0023 was found in 94% of the re-samples. Likewise, the primary run defined MOTU0004, but the re-sample data indicated that this merged with MOTU0015 in 92% of the re-samples (Table 2.2).

Although the re-sample data supported this loss of two MOTUs, they supported splits in MOTU0008 and MOTU0002 to generate another two MOTUs, which resulted in a mean of 59 MOTUs. For the most part, the primary run was a good representation of how these sequences may cluster using a cut-off of 2%.

2.3.2 Congruence between 2% cut-off MOTUs and morpho-species

Once taxa had been defined, morphological species designation was reunited with the sequences to investigate the ability of MOTU_define.pl to recover morphologically established taxa. The results from the primary run at each cut-off were used to assess when complete monospecific MOTUs were formed (Table 2.3).

At a cut-off value of 2%, only 65% of species were recovered as complete monospecific MOTUs. For some species, e.g. *Echinothrips americanus*,

Parthenothrips dracaenae and *Thrips minutissimus*, sequences were found in complete monospecific MOTUs. Other species split into multiple monospecific MOTUs. *Thrips vulgatissimus* was split into two MOTUs, MOTU0017 containing 11 sequences and MOTU0055 containing only T63 (Table 2.2) so was not found to be completely monospecific (Table 2.3). The two specimens of *Limothrips denticornis* did not form one MOTU (Table 2.3): the two sequences were always in separate MOTUs (MOTU0034 and MOTU0044, Table 2.2). *T. palmi* and *Thrips flavus* also formed multiple monospecific MOTUs. A 2%, there were two MOTUs which contained more than one morpho-species. *Haplothrips leucanthemi* and *Haplothrips statices* formed a single MOTU. Two other species, *Thrips fuscipennis* and *Thrips sambuci*, also did not form monospecific MOTUs. Even at 0% difference, they coalesced in a single MOTU. Although these five specimens have been morphologically designated as two species, the molecular barcode results suggested they belong to a single taxon.

Both *Scirtothrips dorsalis* and *Sericothrips staphylinus* seemed to 'lose' complete monospecificity at 2% (Table 2.3). However, having investigated the re-sample data (Table 2.2), it was clear that the primary run MOTUs defined for these two species were not robustly supported. Complete monospecific MOTUs were recovered in 81 and 94 re-samples respectively. There were other discrepancies in the monospecificity of other species such as *Thrips nigropilosus* and *Chirothrips manicatus* where MOTUs were completely monospecific for several cut-offs, then appear not to be, but then complete monospecificity was recovered (grey cells, Table 2.3). This is likely to be due to the addition order issue identified earlier (Table 2.2).

The 2% cut-off was not able to define completely monospecific MOTU, for example MOTU0047 and MOTU0020 contained more than one species (Tables 2.2 and 2.3). Thus 2% was not a sufficient cut-off for definition of complete monospecific MOTU across the thrips data. Examination of MOTU across the different cut-offs showed that specimens were clustered into complete monospecific MOTU at different cut-offs. A maximum of 80.4% morpho-species were in complete monospecific MOTU between 4.4 and 6.6% sequence cut-off (Table 2.3). Looking at this range of cut-off values for the mean number of MOTUs defined (Figure 2.1), there was also a plateau.

There were three species that had only one specimen and it was expected they would form complete monospecific MOTUs. *M. floridensis* (sequence T231) formed a complete monospecific MOTU over all the cut-offs. *H. leucanthemi* was only completely monospecific below 0.66% difference (3 bp). At this cut-off, *H. leucanthemi* grouped with *Haplothrips statices* to form a mixed MOTU. *Odontothrips ignobilis* also had one specimen and formed a complete monospecific MOTU up to 7.7% (35 bp) (Table 2.3). *Aeolothrips albicinctus* had three specimens and was the only other species that formed a completely monospecific MOTU over all the cut-offs. Most MOTUs were made up of a single species but at 2% difference, not all MOTUs were completely monospecific. MOTUs which contained more than one species did not contain species from more than one genus.

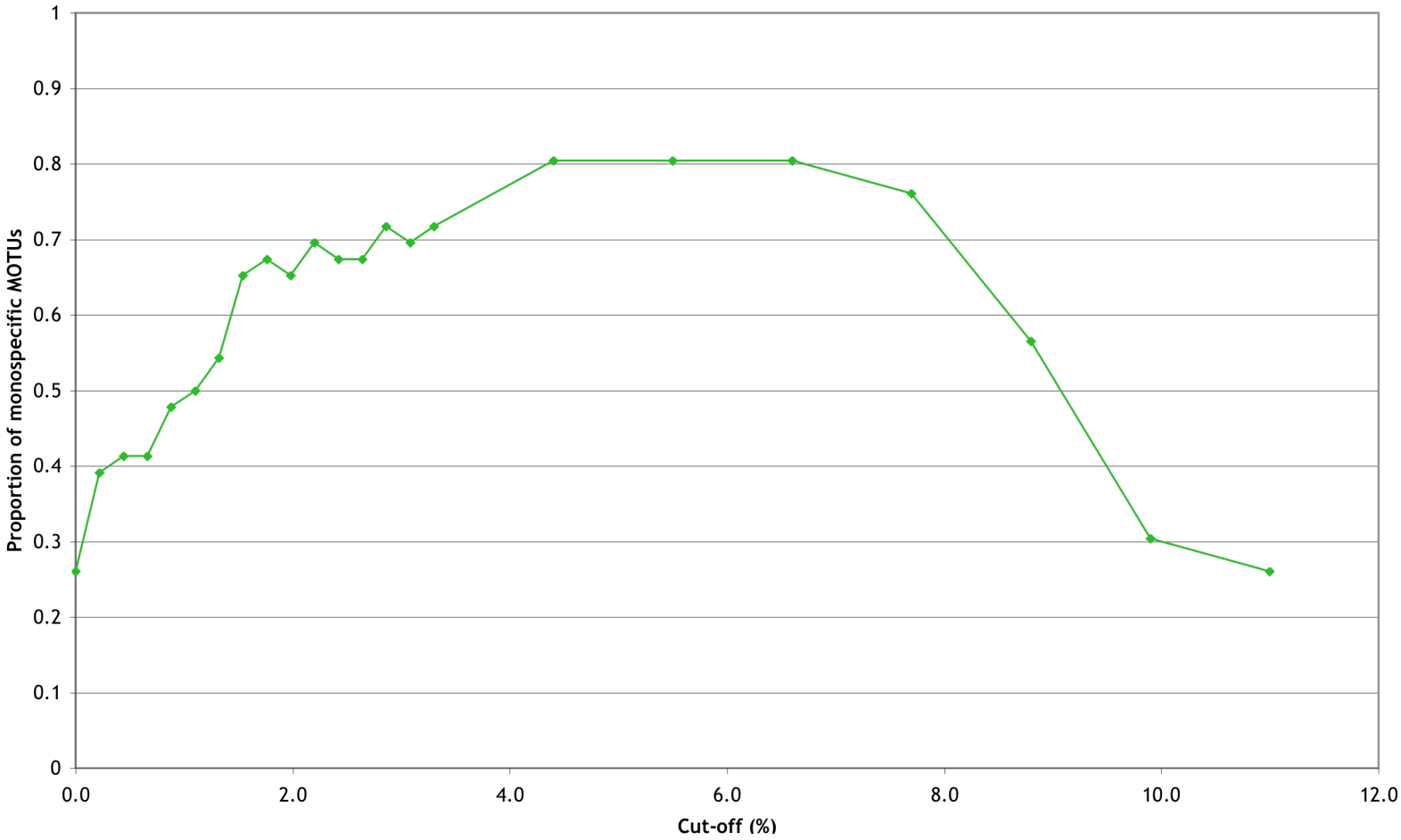
Investigating the proportion of complete monospecific MOTUs over the range of cut-offs could, potentially, highlight a cut-off at which all species were completely monospecific (Figure 2.2). As the cut-off was increased, the proportion of MOTUs that were completely monospecific also increased. There were fluctuations in the proportion as the cut-off was changed from 1.98% to 3.3% probably due to sequence addition order (as only the primary run was investigated). The proportion then remained constant from 4.4% to 6.6% at the maximum of 0.804 (Figure 2.2). Eventually the cut-off value no longer discriminated between species and started to merge multiple species in to a single MOTU, so the proportion of complete monospecificity decreased after 6.6% (30 bp) cut-off. Notably, even the best cut-off (4.4%) still yielded approximately 20% incomplete monospecific MOTU. Two species that have already been identified as never forming monospecific MOTUs, *T. fuscipennis* and *T. sambuci* (Table 2.3), represented only 4.3% of the total species studied. The molecular data thus offer variable support for the morphological taxa. At 0.2% (1 bp) cut-off, *H. leucanthemi* was monospecific, but *T. palmi* was not monospecific until the cut-off was increased to 6.6% (30 bp). These findings suggest a single 'universal' cut-off of 2% does not identify completely monospecific taxa.

Table 2.3 Results from investigating complete monospecificity of primary run MOTUs over a range of cut-offs including the number of sequences per species. If all sequences of a single morpho-species form a single MOTU, then the MOTU scores one. If the morpho-species does not form a single MOTU, then it scores zero. Cut-offs were a percentage of the mean sequence length. Complete monospecificity of morpho-species is highlighted in red. Species that never exhibited complete monospecific MOTUs are highlighted in blue. Queried losses of monospecific MOTUs are shaded in grey. * Species *T. trehernei* and *T. trehernei-pelikani* are grouped together as *T. trehernei*.

Species	Cut-off (%)																						
	0.0	0.2	0.4	0.7	0.9	1.1	1.3	1.5	1.8	2.0	2.2	2.4	2.6	2.9	3.1	3.3	4.4	5.5	6.6	7.7	8.8	9.9	11
A. albicinctus	1 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cep. monilicornis		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Cer. ericae	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ch. manicatus	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0
Ch. meridionalis	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
D. degeeri	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
E. americanus	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
F. intonsa	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
F. occidentalis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
F. schultzei	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
Ha. aculeatus	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Ha. cenchricola	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
H. distinguendus	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Ha. leucanthemi	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ha. setiger	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Ha. statices	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ha. subtilissimus	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
He. femoralis	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Hy. adolfifriderici	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
K. robustus	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
L. denticornis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0
M. floridensis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Od. biuncus	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Od. ignobilis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Od. loti	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Od. phalaratus	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Od. ulicis	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Ox. ajugae	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	1
P. dracaenae	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
Sc. dorsalis	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0
Se. staphylinus	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
T. angusticeps	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
T. brevicornis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1
T. flavus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0
T. fuscipennis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T. hawaiiensis	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. major	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
T. minutissimus	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
T. nigropilosus	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	1	1	0
T. palmi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
T. sambuci	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T. tabaci	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0
T. trehernei *	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
T. urticae	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
T. validus	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
T. vulgatissimus	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0
Total	12	18	19	19	22	23	25	30	31	30	32	31	31	33	32	33	37	37	37	35	26	14	12
Cut-off	**0.0**	**0.2**	**0.4**	**0.7**	**0.9**	**1.1**	**1.3**	**1.5**	**1.8**	**2.0**	**2.2**	**2.4**	**2.6**	**2.9**	**3.1**	**3.3**	**4.4**	**5.5**	**6.6**	**7.7**	**8.8**	**9.9**	**11**

Figure 2.2 Graph of proportion of MOTUs that were completely monospecific for primary runs at each cut-off. The graph never reaches one; there are two species that never form monospecific MOTUs. Conversely the line never drops to zero, as there are two species that always form complete monospecific MOTUs.



2.3.3 DOTUR results

Generating DNA distance matrices using Kimura "2-Parameter" and Jukes-Cantor models produced only minor differences in the number of DOTUs found (Figure 2.3).

As expected, the three DOTUR clustering methods defined different numbers of DOTUs. Furthest neighbour clustering was the most stringent so defined the most DOTUS. As nearest neighbour clustering will allow sequences to form DOTUs which were less similar, it defined the fewest of DOTUs. Average neighbour clustering (an average of the first two methods) defined an intermediate number of DOTUs (Figure 2.3).

DOTUR results from Jukes-Cantor distance matrices using nearest and furthest neighbour clustering were compared with MOTU_define.pl results (Figure 2.4). As expected, MOTU_define.pl and nearest neighbour clustering designated similar numbers of OTUs, but there were consistently fewer MOTUs than nearest neighbour DOTUs (nn-DOTUs) (Figure 2.4). Furthest neighbour clustering (fn-DOTUs) produced more OTUs than the previous two methods. The number of MOTUs was similar to the number of nn-DOTUs until the cut-off reached 7.6% where there was a sudden decrease in the number of MOTUs. There was a steady decrease in the number of DOTUs found for both clustering methods.

At 2% difference, 65 nn-DOTUs were defined. Furthest neighbour clustering created simple splits in nearest neighbour DOTUs to define 81. There were no cases of complex rearranging of sequences in the DOTUR results. When MOTUs were compared with nn-DOTUs, the results were broadly similar. Of the 59 MOTUs found, 44 had equivalent nn-DOTUs and fn-DOTUs (Figure 2.5.A). There were 15 MOTUs which did not have equivalent DOTUs (Figure 2.5.B). Nearest neighbour clustering split seven MOTUs into 16 nn-DOTUs (blue lines, Figure 2.5.B). However, nearest neighbour clustering merged six MOTUs into 3 nn-DOTUs (Figure 2.5.B).

Figure 2.3 Number of DOTUs generated using Jukes-Cantor (JC) and Kimura "2-Parameter" (K2P) distance matrices for all three clustering methods (furthest, nearest and average neighbour).

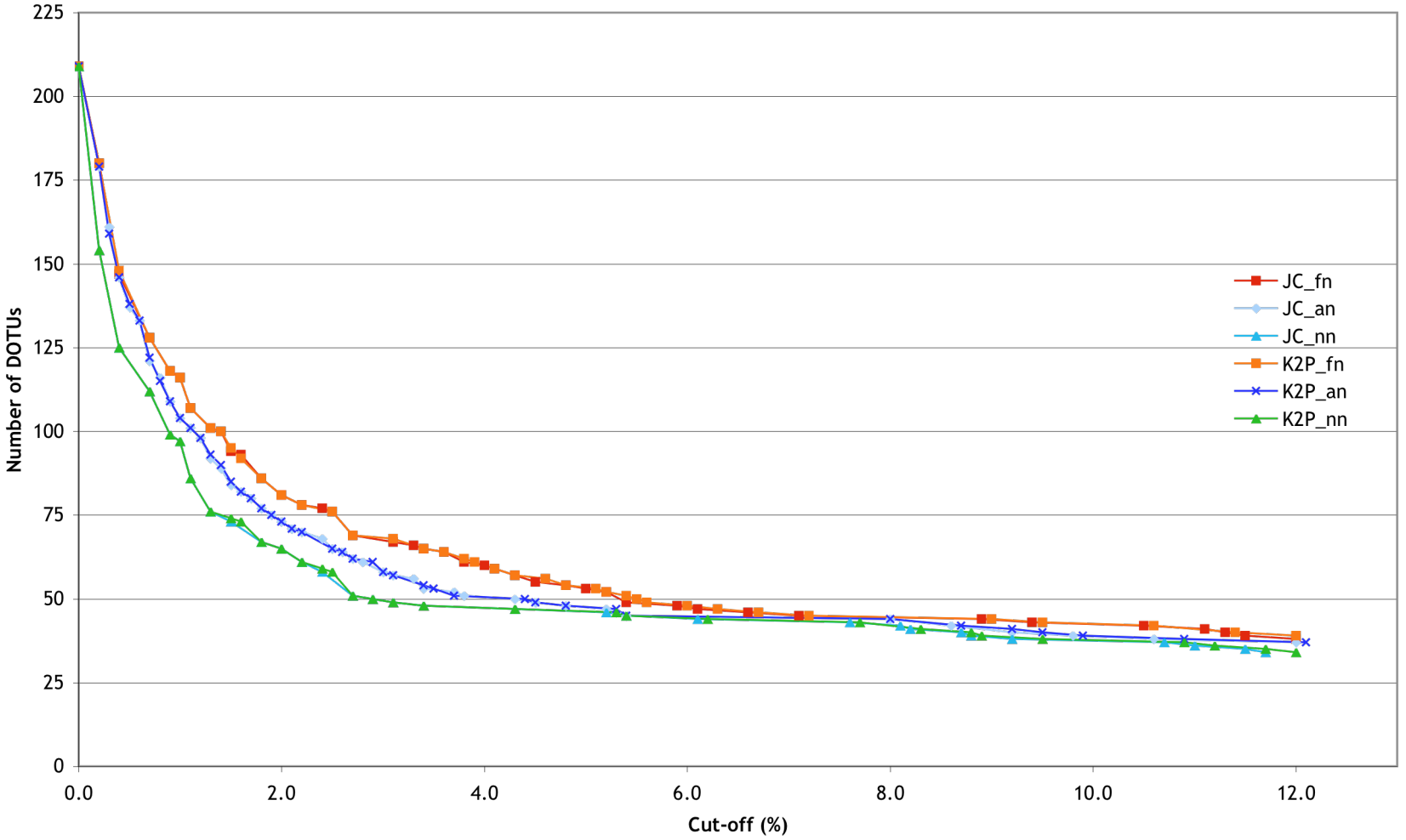


Figure 2.4 Comparison of the number of DOTUs (using matrices with Jukes-Cantor rates (JC) and nearest (nn) and furthest (fn) neighbour clustering) and mean number of MOTUs with standard deviations, designated at each cut-off as a percentage of the mean sequence length.

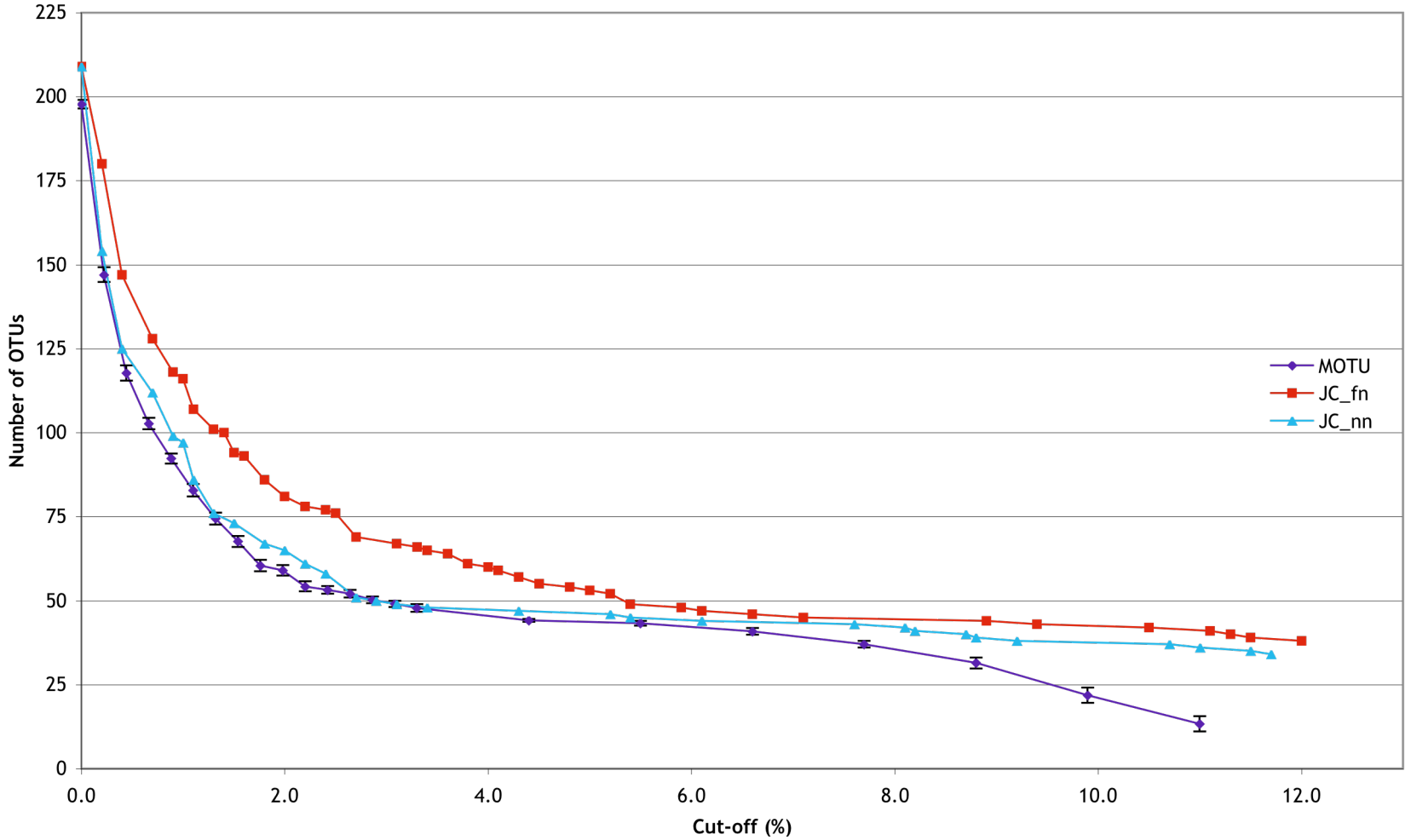


Figure 2.5 Comparison of MOTUs and DOTUs at 2% difference. MOTU_define.pl designation, 59 MOTUs from the primary run at 9 b, was used to compare the members of DOTUs using nearest and furthest neighbour clustering methods. Shaded boxes indicate MOTU_define.pl designation, circles indicate nearest neighbour DOTUs and rectangles indicate furthest neighbour DOTUs. **Figure 2.5.A** shows 44 MOTUs that were equivalent for furthest and nearest neighbour clustering methods. Numbers indicate the number of sequences in the OTU.

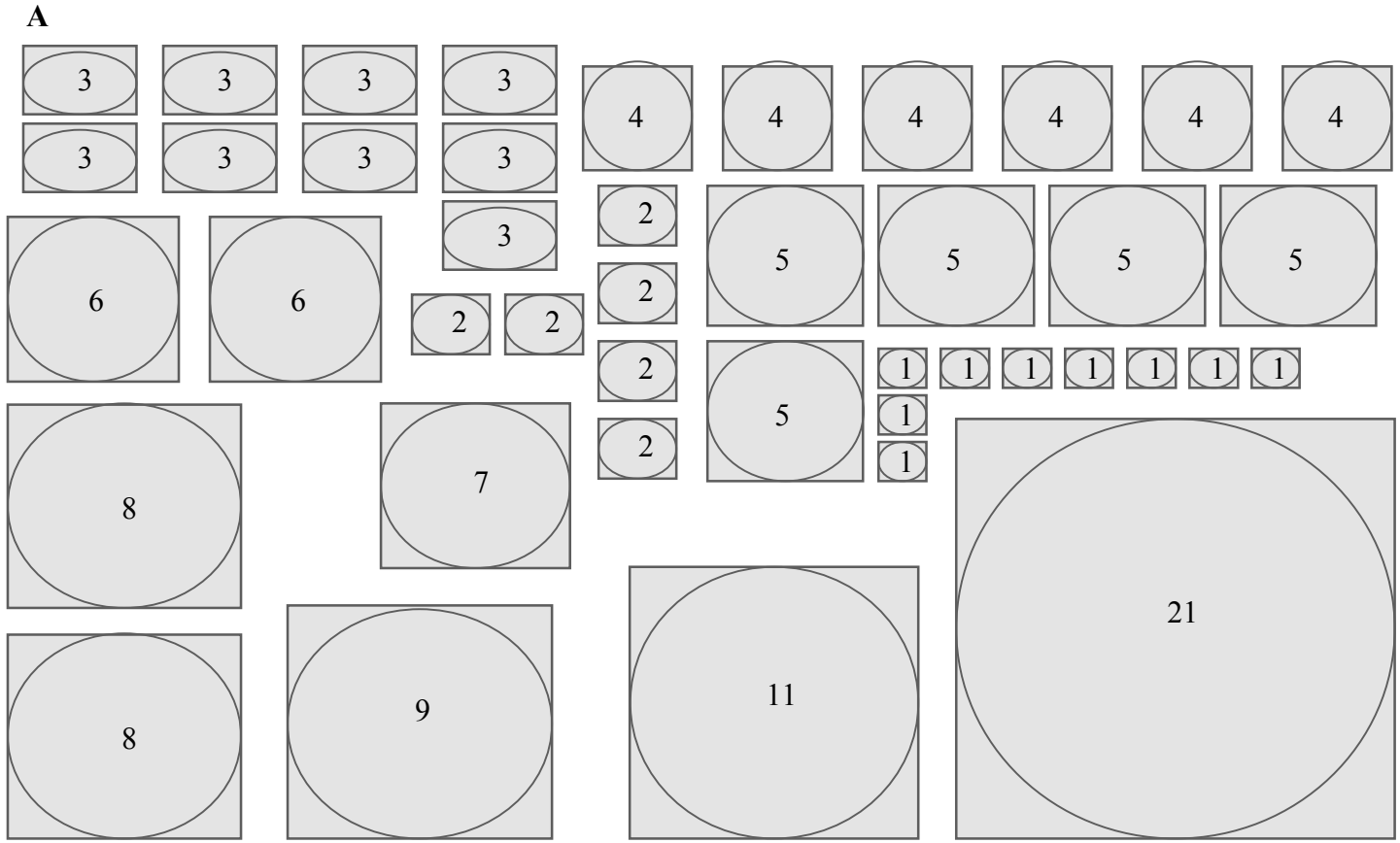
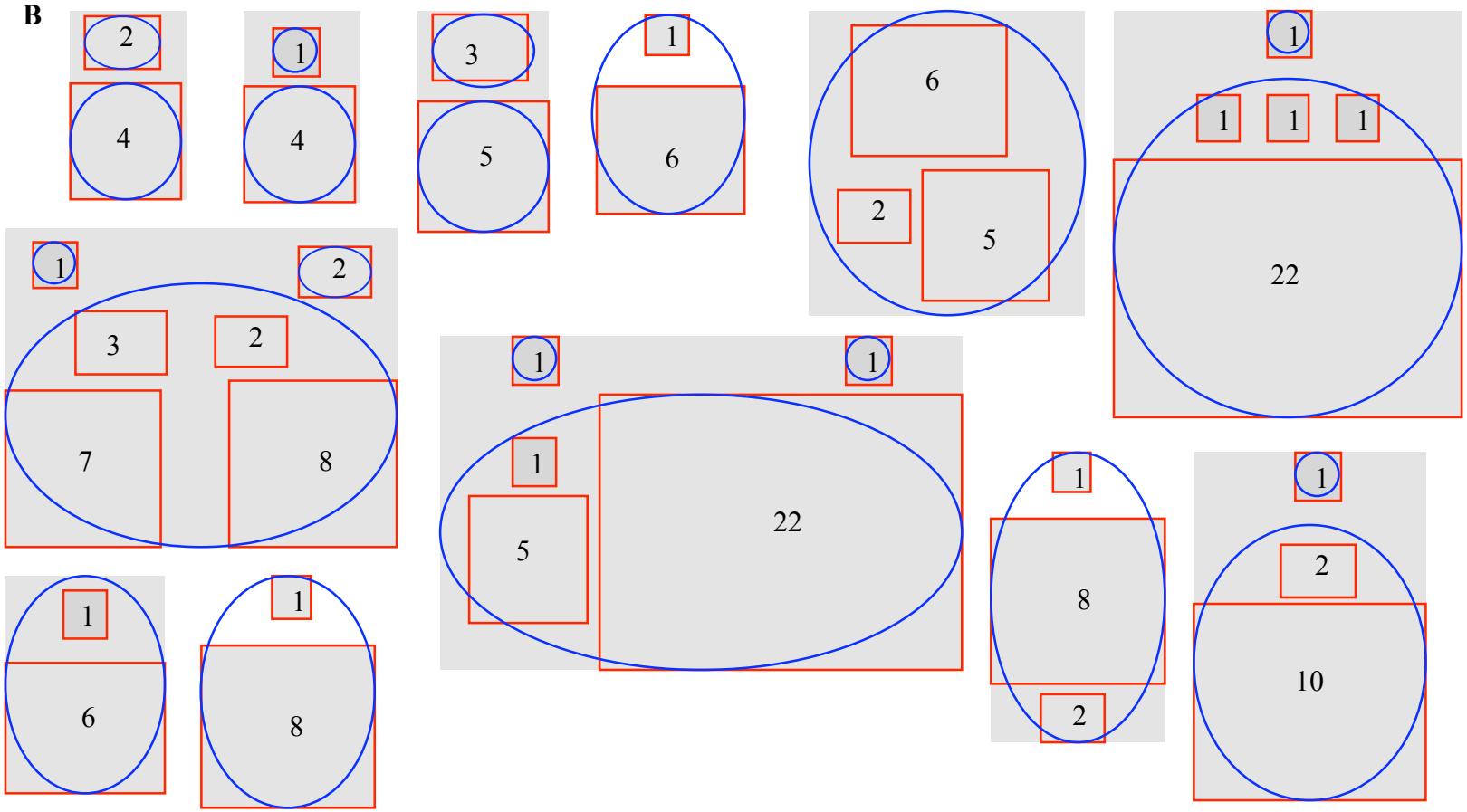


Figure 2.5.B Shows how the remaining 15 MOTUs differ for furthest and nearest neighbour clustering. Shaded boxes indicate MOTU designation, blue lines indicate nearest neighbour clustering DOTUs and red lines indicate furthest neighbour clustering DOTUs. Numbers indicate the number of sequences in OTUs.



Interestingly, these six MOTUs had equivalent fn-DOTUs. There were no joining events when comparing MOTUs and fn-DOTUs. All MOTUs that were split to form nn-DOTUs were also split to form fn-DOTUs, although in some cases there were multiple splits (blue lines compared with red lines, Figure 2.5.B).

2.3.4 Congruence between DOTU and morpho-species

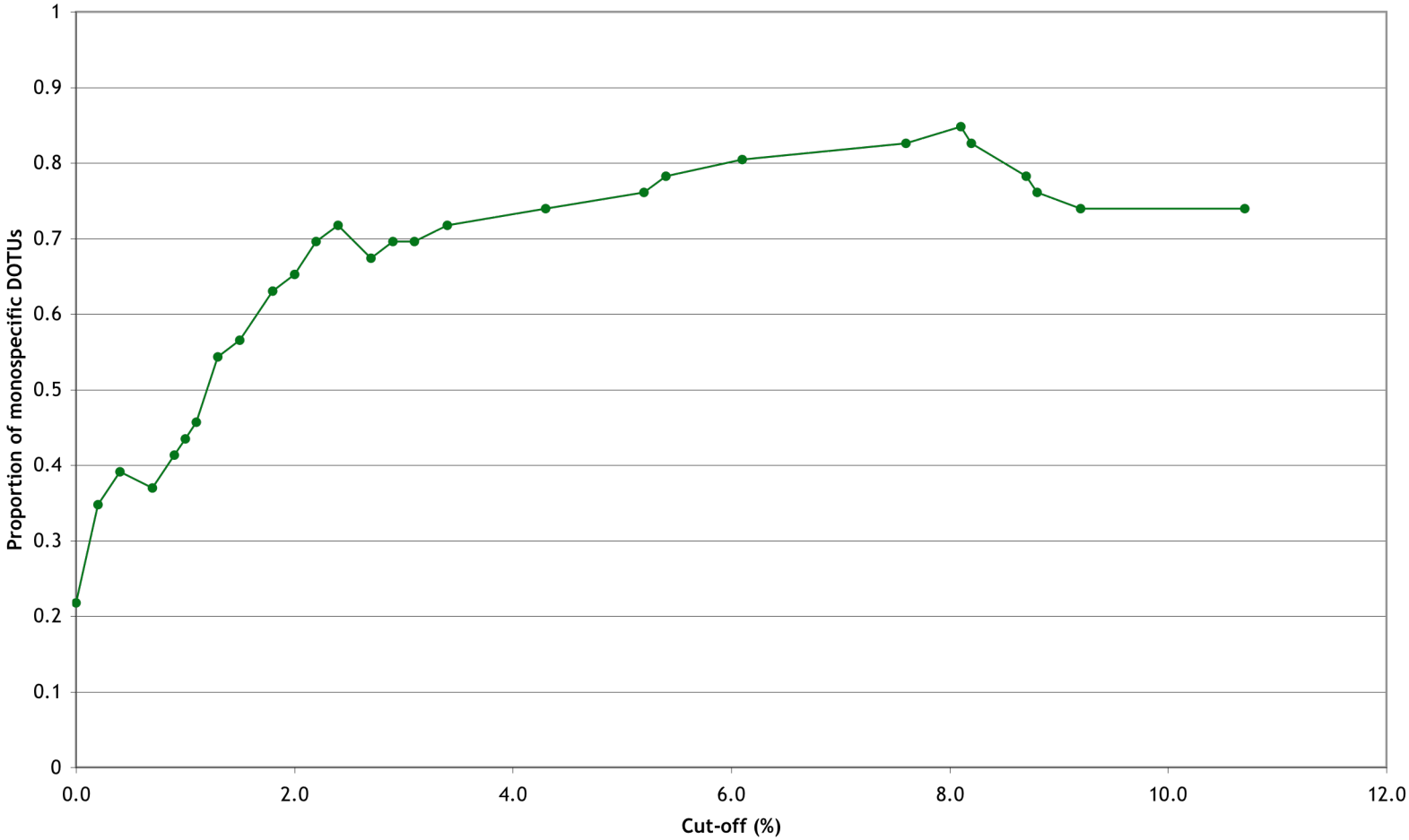
As with the MOTU results, the ability of DOTUR (using nearest neighbour clustering) to produce monospecific DOTUs was also investigated (Table 2.4). *T. fuscipennis* and *T. sambuci* (species that did not form monospecific MOTUs), did not form monospecific DOTUs (highlighted in blue, Table 2.4). DOTUR also found *A. albicinctus* and *M. floridensis* to be monospecific over the range of cut-offs reported. Moreover, DOTUR found three additional species to be monospecific at all cut-off values (highlighted in red, Table 2.4). The pattern of monospecific DOTUs was broadly similar to those found by MOTU_define.pl. There were no losses and recoveries of monospecificity at particular cut-offs as seen in the MOTU_define.pl results.
The proportion of monospecific DOTUs was also examined (Figure 2.6). Between cut-off values of 0.0% and 0.4%, the proportion of monospecific DOTUs increased to approximately 0.4. There was then a drop in the proportion before it began to rise to 0.72 when the cut-off value was 2.4% (Figure 2.6). There was then another fall before the proportion reached a maximum of 0.85 at a cut-off value of 8.1%. The proportion then fell again, before levelling off at 0.74 at between cut-offs 9.2% and 10.7%. As there was no effect of addition order on the way DOTUR performs sequence analysis, there is no clear explanation for the decreases in proportion of monospecificity. The maximum proportion of monospecific DOTUs (0.847) was higher than monospecific MOTU (0.804). However this was reached at a cut-off value of 8.1% for DOTUs compared with 4.4 - 6.6% for MOTUs.

Table 2.4 Monospecific groupings from DOTUR (Jukes-Cantor distances, nearest neighbour clustering) results. Colour coding and * as in Table 2.3.

Species	Cut-off																											
	0.0	0.2	0.4	0.7	0.9	1.0	1.1	1.3	1.5	1.8	2.0	2.2	2.4	2.7	2.9	3.1	3.4	4.3	5.2	5.4	6.1	7.6	8.1	8.2	8.7	8.8	9.2	10.7
A. albicinctus	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cep. monilicornis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cer. ericae	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ch. manicatus	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ch. meridionalis	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
D. degeeri	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
E. americanus	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F. intonsa	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
F. occidentalis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
F. schultzei	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ha. aculeatus	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
Ha. cenchricola	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
H. distinguendus	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
Ha. leucanthemi	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ha. setiger	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ha. statices	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ha. subtilissimus	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
He. femoralis	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Hy. adolfifriderici	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
K. robustus	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
L. denticornis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
M. floridensis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Od. biuncus	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Od. ignobilis	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Od. loti	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Od. phalaratus	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Od. ulicis	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
Ox. ajugae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
P. dracaenae	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Sc. dorsalis	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Se. staphylinus	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. angusticeps	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

T. brevicornis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
T. flavus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
T. fuscipennis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T. hawaiiensis	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. major	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. minutissimus	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. nigropilosus	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. palmi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
T. sambuci	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T. tabaci	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
*T. trehernei**	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. urticae	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. validus	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
T. vulgatissimus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
Total	10	16	18	17	19	20	21	25	26	29	30	32	33	31	32	32	33	34	35	36	37	38	39	38	36	35	34	34
Cut-off	**0.0**	**0.2**	**0.4**	**0.7**	**0.9**	**1.0**	**1.1**	**1.3**	**1.5**	**1.8**	**2.0**	**2.2**	**2.4**	**2.7**	**2.9**	**3.1**	**3.4**	**4.3**	**5.2**	**5.4**	**6.1**	**7.6**	**8.1**	**8.2**	**8.7**	**8.8**	**9.2**	**10.7**

Figure 2.6 Graph of proportion of DOTUs that were monospecific at reported distance levels. Points on the graph did not achieve the upper limit (1); two species did not form monospecific DOTUs. Conversely the line never drops to zero, as six species consistently formed monospecific DOTUs.

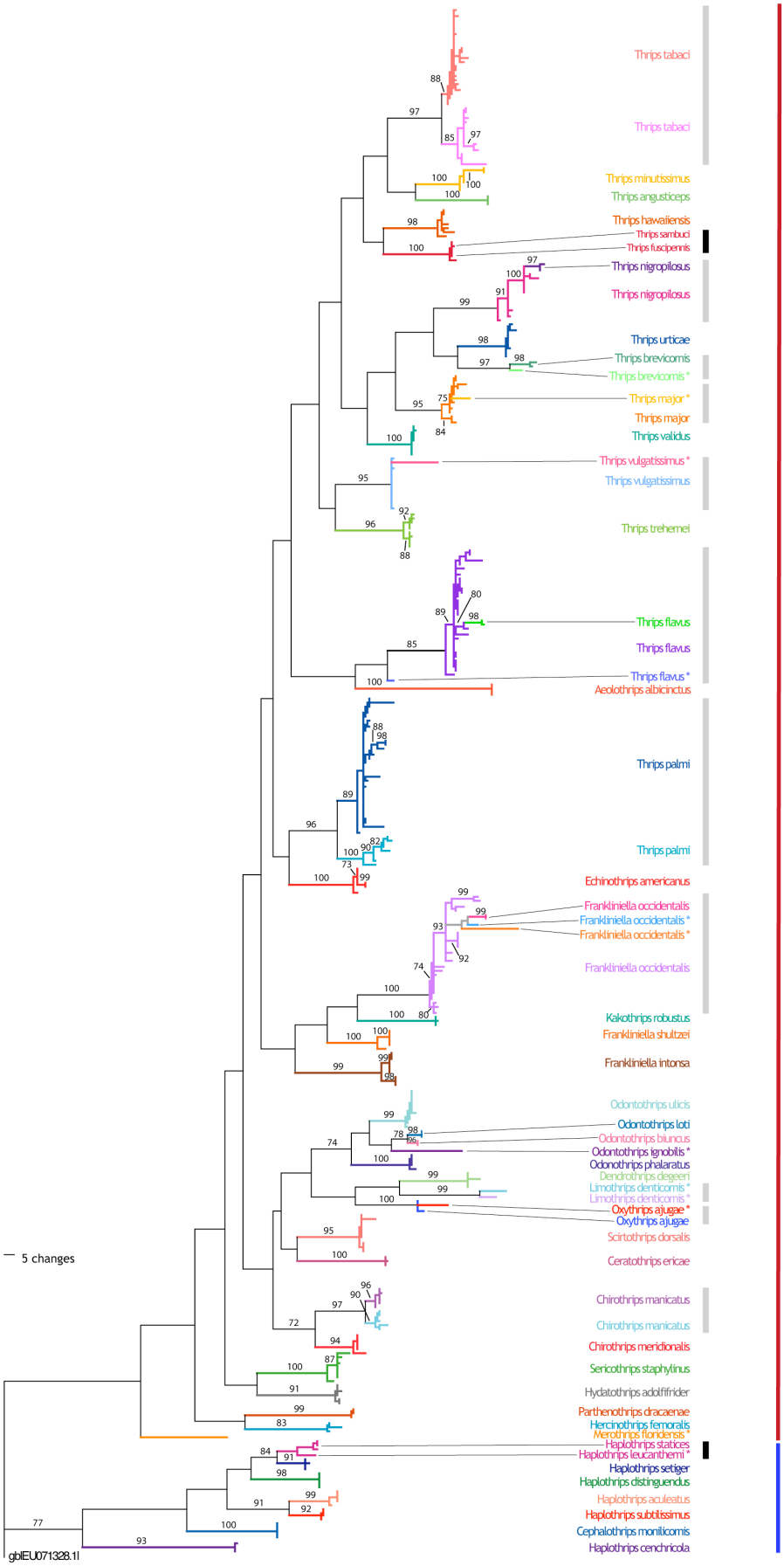


Moreover there was no plateau in the proportion of monospecific DOTUs, so highlighting a phase where the number of DOTUs was stable was problematic.

2.3.5 Phylogenetic comparison of Thrips data to OTUs

The ML tree generally supported OTUs defined by MOTU_define.pl and DOTUR (Figure 2.7). Some morpho-species, such as *Aeolothrips albicinctus* and *Cephalothrips monilicornis*, were strongly supported and were deeply branched. These were also species that formed complete monospecific OTUs. Others had strongly supported divisions within morpho-species (e.g. *T. tabaci* and *T. palmi*) but the lineages were sister taxa. These splits were also found in MOTUs and DOTUs. Not all genera were monophyletic. The genera *Frankliniella* had *Kakothrips* nested within the clade, and *Cephalothrips* was nested in the *Haplothrips* lineages. The data also supported the segregation of Terebrantia and Tubulifera.

Figure 2.7 Phylogram of thrips data set. Bootstrap values of over 70 (from 2000 replicates) are reported. Different colours indicate different MOTUs defined using a 9 bp cut-off value. Black bars indicate multi-specific MOTUs, grey bars indicate multiple MOTUs defined for a single morpho-species. The two sub-orders are also shown, Tubulifera (blue bar) and Terebrantia (red bar).


Thrips tabaci
Thrips tabaci
Thrips minutissimus
Thrips angusticeps
Thrips hawaiiensis
Thrips sambuci
Thrips fuscipennis
Thrips nigropilosus
Thrips nigropilosus
Thrips urticae
Thrips brevicornis
Thrips brevicornis *
Thrips major *
Thrips major
Thrips validus
Thrips vulgatissimus *
Thrips vulgatissimus
Thrips trehernei
Thrips flavus
Thrips flavus
Thrips flavus *
Aeolothrips albicinctus
Thrips palmi
Thrips palmi
Echinothrips americanus
Frankliniella occidentalis
Frankliniella occidentalis *
Frankliniella occidentalis *
Frankliniella occidentalis
Kakothrips robustus
Frankliniella shultzei
Frankliniella intonsa
Odontothrips ulicis
Odontothrips loti
Odontothrips biuncus
Odontothrips ignobilis *
Odonothrips phalaratus
Dendrothrips degeeri
Limothrips denticornis *
Limothrips denticornis *
Oxythrips ajugae *
Oxythrips ajugae
Scirtothrips dorsalis
Ceratothrips ericae
Chirothrips manicatus
Chirothrips manicatus
Chirothrips meridionalis
Sericothrips staphylinus
Hydatothrips adolfifrider
Parthenothrips dracaenae
Hercinothrips femoralis
Merothrips floridensis *
Haplothrips statices
Haplothrips leucanthemi *
Haplothrips setiger
Haplothrips distinguendus
Haplothrips aculeatus
Haplothrips subtilissimus
Cephalothrips monilicornis
Haplothrips cenchricola
5 changes
gbIEU071328.1I
Terebrantia
Tubulifera

2.4 Discussion

2.4.1 MOTU_define.pl

MOTU_define.pl clusters sequences based on similarity from pair-wise comparisons. As the cut-off increases, the probability that a sequence will match any member of a MOTU increases, and the number of MOTUs declines. Initially there was a sharp drop in the number of MOTU defined, but as the cut-off increased, the fall in the number of MOTU decreased (Figure 2.1). MOTU_define.pl generated a plateau in the number of MOTU defined over a range of cut-offs, from 4.4% to 6.6%, where the mean number of MOTU dropped from 44 to 43 (Figure 2.1). Importantly, this plateau is expected to indicate cut-offs at which the number (and members of) MOTUs were stable, i.e. a barcoding gap. Previous work by Morris and Mound (2001) shows that intraspecific distances of some thrips species can be higher than this. This barcoding gap is only visible if a range of cut-off values is used to investigate the clustering behaviour of sequences. Re-sampling the data set allowed the robustness of those clusters to be verified (Table 2.2). Using MOTU_define.pl, and using a range of cut-offs and multiple re-samplings, allows a user to interrogate the data fully.

2.4.2 Monospecificity of MOTUs

MOTU_define.pl is agnostic of any species designation. When species identifications were reunited with sequences, MOTUs defined tended to be made up of only one species. Whilst the maximum proportion of complete monospecific MOTUs was attained at the same values as the barcoding gap, it did not reach 1. Although the difficulties in identifying small organisms are well documented, it is worthwhile to note that there were only two species that never form monospecific molecular taxa (Table 2.3). The ability of traditional taxonomists to define species based on morphological characteristics that has a high level of concordance with molecular taxa must be acknowledged. However, this dataset consists of only a small proportion of all thrips species and is not an exhaustive global sample.

MOTU_define.pl can be applied to any level of a phylogenetic tree depending on the cut-off used. A low cut-off value may be suitable for identifying species although it may be confounded by intraspecific variation. However, genera within families are more difficult to delimit as they can be of different ages. Older genera will be segregated more easily than younger genera which are closely related and monophyly would be reached at different cut-off values.

2.4.3 DOTUR

DOTUR uses a DNA distance matrix generated from a multiple alignment of sequences to report the distance at which the number of DOTUs changes. Consistently, the number of DOTUs reduced with increasing distance and no clear plateau was formed (Figure 2.3) whichever distance or clustering method was used. There is no objective way to reveal a barcoding gap. Unless a suitable cut-off has previously been defined then DOTUR is less suitable for taxa identification than MOTU_define.pl.

There were very minor differences between the Jukes-Cantor and Kimura "2-Parameter" results (Figure 2.3). The Kimura-"2-Parameter" models two rates of evolution based on the dataset. When genetic distances within the dataset are low, these two rates are approximately equal, so the model approaches the Jukes-Cantor model. When comparing the same gene across organisms, this is likely to be true as orthologous sequences are subject to similar constraints and therefore likely to be highly conserved. If this limits the phylogenetic usefulness of COI sequences, performing analysis based on complex models of evolution becomes redundant. Therefore it is valid to use the Jukes-Cantor single-rate model in comparison with MOTU_define.pl results.

Even though using furthest neighbour clustering is likely to overestimate the number of DOTUs in the data set, nevertheless this particular method is currently the most stringent as it requires all sequences in the DOTU to be within the similarity cut-off. Nearest neighbour clustering is more conservative in the estimation of DOTUs. This method defines OTUs using

sequence similarity to any of the sequences in a DOTU. This is the most similar to the way MOTUs are defined.

2.4.4 Monospecificity of DOTUs

DOTUR did not fully sort sequences into monospecific taxa. As with MOTU results, *T. fuscipennis* and *T. sambuci* were not monospecific in DOTUR over the range of cut-offs investigated. However, these two species do not account for the missing monospecific taxa. The peak proportion of monospecific DOTU found was higher than for MOTU. DOTUs are not influenced by addition order (unlike MOTUs) so the fluctuations (at 0.7% and 2.7%, Figure 2.6) in the proportion of monospecific nn-DOTUs are not easily explained.

2.4.5 MOTU vs. DOTUR

The main difference between MOTU_define.pl and DOTUR is the method used to designate taxa. MOTU_define.pl uses BLAST to form pairwise comparisons between all sequences in the data set. DOTUR requires a full alignment of all the sequences, followed by a derived distance matrix, which uses two models of evolution to the data, first in the multiple-alignment and secondly in the construction of the DNA matrix. It may be assumed that these models accurately reflect the evolutionary history of the sequences. If the alignment is poor at the ends (where sequence quality can be lower) or sequences are of different lengths, DOTUR interprets these as real and generates DOTUs accordingly. For this data set the sequences were of varying lengths and this may explain the higher numbers of DOTUs versus MOTUs. Even when DOTUs were generated using nearest neighbour clustering, there were still more DOTUs. It is likely that using DOTUR for species identification from barcodes will overestimate the number of taxa, where as MOTU_define.pl will be more conservative.

In the case of COI barcodes, it may be inappropriate to investigate taxon composition using methods driven by an evolutionary hypothesis. Although they can easily identify "species groups", it is not possible to resolve deep

phylogeny with these sequences (Meyer and Paulay, 2005). DOTUR models the nodes of the phylogeny. The problems with using COI as a barcode are well reported. Not only is it unreliable to extract, the ability to resolve closely related species is under scrutiny (Hurst and Jiggins, 2005). When using barcodes for species identification, the sequence in question needs only to be compared to an identified sequence and asked does it match a known sample or not? Species delimination from sequence is a more complex process as the differences between sequences are considered to have some sort of biological significance that relates to the difference between species, such as a difference in colour. In this case, it would be prudent to incorporate models of evolution into analysis where the hypothesis being tested is how one species relates to another.

Both MOTU_define.pl and DOTUR grouped *T. fuscipennis* and *T. sambuci* in the same OTUs and neither found them to be monospecific. Although morphologically these have been designated has separate species, the molecular data suggested they belong a single OTU. The defining feature between the two species is antennae colouration, which is the final couplet of the identification key (D. Collins, personal communication). The colouration (or lack of it) may separate the two species morphologically, but does not show any molecular difference when looking at COI barcodes.

2.4.6 Is a cut-off of 2% sufficient to discriminate thrips species?

Some proponents of barcoding have suggested that 2% difference of COI sequences will be sufficient and reliable to discriminate species (Hebert *et al.*, 2003b). At this cut-off, all the sequences from a single morpho-species should be in a complete monospecific OTU. In addition, all OTUs should be monospecific, containing only one morpho-species. However, both DOTUR and MOTU_define.pl found maximal proportions of complete monospecific OTUs at higher cut-offs, 4.4% and 8.1% respectively. These results for interspecific divergences are lower than previous work on thrips COI sequences. Mound and Morris (2001) found that intraspecific differences were 6.1% and 8.4% for COI sequences of two thrips species, and divergence between the species was 14%. If these any of these divergence values hold for

the majority of thrips species, then using 2% will illustrate variation associated with populations as species boundaries.
It would be unrealistic to expect that any one cut-off would generate OTUs which were completely consistent with morpho-species.

2.4.7 Significance of barcodes for thrips

The sub-orders Terebrantia and Tubulifera were segregated by phylogenetic analysis (Figure 3.7). MOTU_define.pl and DOTUR also separated the two sub-orders. Within the Tubulifera, *H. statices* and *H. leucanthemi* were defined as belonging to the same OTU for the majority of cut-offs used. Other morpho-species with in the Tubulifera were well identified by OTUs as they tended to be completely monospecific for most of the cut-offs used.
There were also morpho-species within the Terebrantia that were well identified such as *Kakothrips robustus* and *Ceratothrips ericae*. There is one example where two taxa are not clearly recognized. The five *T. sambuci* and *T. fuscipennis* sequences always intermingle, whether as OTUs or in a phylogenetic tree. This would indicate that these species designations need revision as the morphological feature, or features, used to currently distinguish the two, are not represented in the molecular data.
There are several morpho-species which split into multiple OTUs and the corresponding clades have well supported branches on the phylogenetic tree. Two clusters of *Chirothrips manicatus* are seen in OTU results and in the tree. *Thrips palmi* and *T. tabaci* also show a deep split in the data set. It is unclear what the cause of these splits are. Populations from different geographic locations which are ecologically similar may maintain morphology but would experience genetic drift within the populations. Examining the *T. tabaci* clades, shows no simple correlation between MOTU (Table 2.2) or clades (Figure 2.7) and geography (Appendix 2.2). Specimens from the same location cluster together, but not all specimens from the UK are in the same clade. Alternatively, the divisions may be indicative of cryptic or incipient speciation.
Whatever the cause, such cases of major partitioning of specimens warrant further investigations to reconnect morphological and molecular data. When

the taxon is of major agricultural interest, as *T. tabaci* is, the validity of a species is particularly important.

Whilst this study is by no means an exhaustive representation of thrips molecular diversity, using MOTUs to define taxa is a promising approach for thrips surveillance. There are issues regarding the current taxonomic status of some species which may benefit from further investigation. Increasing the breadth and depth of sampling of species, especially those which only had a one or two specimens (e.g. *M. floridensis* and *L. denticornis*) and from the sub-order Tubulifera, would also increase the value of the data set. MOTU_define.pl can define thrips taxa from this data set and in the long term should provide identification of samples quicker than traditional morphological methods.

Appendix 2.1

As part of my major undergraduate project, I carried out a molecular survey of tardigrades which has been included in the following paper.

Appendix 2.2

Geographic locations of *T. tabaci* specimens, MOTU07 is the red clade and MOTU35 is the pink clade in Figure 2.7. Specimen numbers in italics come from unknown locations.

MOTU ID	Specimen Number	Location
MOTU07	T97	Neum, Bosnia-Herzegovina
	T124	Neum, Bosnia-Herzegovina
	T39	UK
	T121	Neum, Bosnia-Herzegovina
	T119	Northampton, UK
	T322	Bucks, UK
	T321	Bucks, UK
	T25	Colchester, Essex, UK
	T224	Humberside, UK
	T99	Neum, Bosnia-Herzegovina
	T297	Leigh-on-Sea, Essex, UK
	T162	*Unknown*
	T161	*Unknown*
	T223	Humberside, UK
	T123	Neum, Bosnia-Herzegovina
	T113	*Unknown*
	T222	Humberside, UK
	T298	Leigh-on-Sea, Essex, UK
	T100	Hampshire, UK
	T98	Neum, Bosnia-Herzegovina
	T296	Leigh-on-Sea, Essex, UK
MOTU35	T283	ex. Israel
	T160	*Unknown*
	T284	ex. Israel
	T183	*Unknown*
	T285	ex. Israel
	T180	*Unknown*
	T171	*Unknown*
	T173	*Unknown*
	T172	*Unknown*
	T352	Isles of Scilly, UK
	T353	Isles of Scilly, UK
	T182	*Unknown*
	T354	Isles of Scilly, UK.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2005) **360**, 1935–1943
doi:10.1098/rstb.2005.1725
Published online 12 September 2005

Defining operational taxonomic units using DNA barcode data

Mark Blaxter*, Jenna Mann, Tom Chapman, Fran Thomas, Claire Whitton, Robin Floyd† and Eyualem Abebe‡

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Laboratories, King's Buildings, Edinburgh EH9 3JT, UK

The scale of diversity of life on this planet is a significant challenge for any scientific programme hoping to produce a complete catalogue, whatever means is used. For DNA barcoding studies, this difficulty is compounded by the realization that any chosen barcode sequence is not the gene 'for' speciation and that taxa have evolutionary histories. How are we to disentangle the confounding effects of reticulate population genetic processes? Using the DNA barcode data from meiofaunal surveys, here we discuss the benefits of treating the taxa defined by barcodes without reference to their correspondence to 'species', and suggest that using this non-idealist approach facilitates access to taxon groups that are not accessible to other methods of enumeration and classification. Major issues remain, in particular the methodologies for taxon discrimination in DNA barcode data.

Keywords: DNA barcodes; molecular operational taxonomic units; tardigrades; nematodes; meiofauna; small subunit ribosomal RNA

1. INTRODUCTION: THE UNSEEABLE ANIMAL

The total number of unique taxa described to the species level is circa 1.5 million, but the total number of 'species' is likely to be in the region of 10 million (May 1988). The overall 'taxonomic deficit' (the ratio of expected taxa to named taxa) is thus approximately sixfold. However this deficit, like all phylogenetic things, is not immune to systematic bias (Blaxter 2003). For vertebrates, the current described species total is likely to be relatively close to the 'true' total: we have described most of these relatively large organisms. The same is true of most groups whose members have body sizes greater than 10 mm. However, the vast majority of organisms on the Earth have body sizes less than 1 mm, and for these groups the taxonomic deficit is likely to be several fold worse than for land plants and vertebrates (Lambshead 1993; Platt 1994; Lambshead & Boucher 2003). These meio- and micro-fauna and flora are, however, key to the functioning of ecosystems and are the productive and saprophytic base upon which the macro-organisms rely. Their size precludes facile visual identification, and indeed much of their important morphology may be at scales that are beyond the resolution of light microscopy (De Ley & Bert 2001; De Ley *et al.* 2005). Wendell Berry quotes from his daughter in his poem 'To the unseeable animal': *'I hope there's an animal somewhere that nobody has ever seen./ And I hope nobody ever sees it.'* (Berry 1970). We suggest that DNA barcoding may permit rational access to these animals.

DNA barcoding, the use of a specified DNA sequence to provide taxonomic identification for a specimen, is a technique that should be applicable to all cellular (and much viral) life (Floyd *et al.* 2002; Hebert *et al.* 2003; Tautz *et al.* 2003; Blaxter *et al.* 2004). Theoretically, this should allow rapid and high-throughput identification, either of individual organisms or of sequences isolated from an environmental DNA sample. Specimen-independent DNA surveys are already used for microbial (Giovannoni *et al.* 1990) and protozoal communities (Diez *et al.* 2001; Lopez-Garcia *et al.* 2001; Moreira & Lopez-Garcia 2002; Amaral Zettler *et al.* 2002), and have revealed a wealth of hidden diversity. Meiofauna would appear to be an ideal group in which a molecular identification system could be used (Lambshead 1993; Lawton *et al.* 1998; Blaxter 2004).

2. BARCODING MEIOFAUNA: CHALLENGES

The number of meiofaunal taxa, animals with a body size ~1 mm (or less), can only be guessed at. Thus, the number of described species of nematodes is quoted as between 26 000 and 40 000, but the real total estimated to be above one million (Lambshead 1993; Platt 1994; Lambshead & Boucher 2003). The deficit may be put into perspective by considering that the number of described species of soil dwelling nematodes for the UK is approximately 400, a figure surprisingly close to the inventory of UK breeding birds. Is the UK nematode fauna really that depauperate? Our surveys of nematodes in soils in relatively degraded habitats (upland farm grassland) suggest that taxon numbers identifiable from even a small area may be remarkably high (R. Floyd, A. Eyualem and M. Blaxter,

* Author for correspondence (mark.blaxter@ed.ac.uk).
† Present address: British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK.
‡ Present address: Department of Zoology, Hubbard Center for Genome Studies, University of New Hampshire, Durham NH 03824, USA.

One contribution of 18 to a Theme Issue 'DNA barcoding of life'.

© 2005 The Royal Society

unpublished). Similarly, for tardigrades, the described UK fauna is ~100 species (Maucci 1986; Kinchin 1994), but we have identified over 50 taxa from one restricted set of sample sites (Blaxter *et al.* 2003). While some authors have argued for a relatively low number of meio-taxa matched by a near-ubiquitous distribution (Finlay 2002), we have found that different sites, though close geographically, can have very different taxon assemblages (Blaxter *et al.* 2003). If organisms with a body size <1 mm really do have no biogeographical structure, and are all essentially ubiquitous, the sampling we have carried out suggests at least that relative abundances must vary greatly between sites. Meiofaunal barcoding must fall into the purview of the third community identified above: experimental investigation of biodiversity.

We have been generating DNA barcode datasets for meiofaunal specimens (mostly nematodes and tardigrades) for several years (Floyd *et al.* 2002; Blaxter & Floyd 2003; Blaxter *et al.* 2003; Eyualem & Blaxter 2003; Blaxter 2004; Blaxter *et al.* 2004). We are agnostic as to whether the taxa we can define using these barcode sequences (which we call 'molecular operational taxonomic units' or MOTU) are 'species' or not, though in the case where we have compared and contrasted MOTU, morphological species hypotheses and breeding-based biological species, MOTU and biological species hypotheses were congruent while morphological analyses disagreed internally, and with the other modes of taxon definition (Eyualem & Blaxter 2003). We have traditionally used the nuclear small subunit (nSSU) as a marker, but have also tested nSSU alongside cytochrome oxidase subunit I (*cox1*), with equivalent resolution. Here, we use a new dataset of meiofaunal barcodes to discuss what we feel are very interesting and important features of DNA barcode data: they can be used not only to define taxa, but also to identify sets of specimens for which robust taxonomic hypotheses are difficult to construct. These clouds of related specimens are immediately of interest for further study: is this evidence for recent, rapid radiation of distinct taxa or is it evidence for a highly variable single taxon?

3. METHODS: OBTAINING MEIOFAUNAL BARCODE SEQUENCES

(a) *Sampling of moss ecosystems*

Moss samples for this study were collected from dry stone walls surrounding Ettrick Old Church, in Glen Ettrick in Southern Scotland (Blaxter *et al.* 2003). Meiofauna were isolated by modified Baermann funnel separation through milk filters into sterile tap water. Larger fauna (such as collembolans and mites; body sizes >2 mm) were excluded from the separation by the pore size of the filter: some of these arthropods were picked from moss individually. Relative numbers of animals from each phylum were counted from a subsample of the filtrate, and a few of each phylum picked individually: the remainder was processed for DNA extraction.

(b) *Individual specimen barcoding*

Individual animals were extracted using the NaOH direct lysis procedure: this yields ~40 µl of stable extract per specimen from which over a dozen PCRs can be performed (Floyd *et al.* 2002). Bulk filtrate animals were concentrated by centrifugation and extracted using a snap-freezing/proteinase K/phenol/chloroform protocol. The nSSU marker was amplified from individual extracts using the primers SSU_F04 (GCTTGTCTCAAAGATTAAGCC) and SSU_R26 (CATTCTTGGCAAATGCTTTCG) (Blaxter *et al.* 1998), yielding a ~900 base pair (bp) product. These primers were designed to be metazoan-specific (Blaxter *et al.* 1998). The *cox1* amplicons were amplified from a subset of tardigrade individuals (also amplified for nSSU) using the 'universal' primers *cox1* (HC02198; TAAACTTCAGGGTGACCAAAAAATCA) and *cox1* (LC01490; GGTCAACAAATCATAAAGATATTGG) (Hebert *et al.* 2003), yielding a ~650 bp product. Shrimp alkaline phosphatase/exonuclease I-cleaned PCR products from single specimens were sequenced directly using SSU_R09 or *cox1*.

(c) *Barcodes from bulk ecosystem DNA*

Bulk filtrate animals were concentrated by centrifugation and extracted using a snap-freezing/proteinase K/phenol/chloroform protocol. The nSSU marker was amplified as described above. Amplicons of nSSU generated from the bulk extract target were cleaned using a Montage gel extraction kit and cloned into pTOPO2.1 (Invitrogen). After growth on LB/kanamycin/IPTG/Xgal, recombinant colonies were picked to 200 µl of LB broth with kanamycin in microtitre plates and grown overnight. Inserts in the recombinant plasmids were amplified from ~1 µl of overnight liquid culture using the primers M13_F (CTGGCCGTCGTTTAC) and M13_R (CAGGAAACAGCTATA), cleaned using shrimp alkaline phosphatase and exonuclease I, and sequenced using SSU_R09 (AGCTGGAATTACCGCGGCTG) and ABI BigDye3.0 reagents to produce ~500 bp of sequence.

(d) *Molecular operational taxonomic unit definition*

The sequencing was carried out on an ABI3730 capillary sequencer, and sequencing chromatograms were post processed with trace2seq (a perl program that uses phred to identify high-quality base calls and crossmatch to identify vector sequence; A. Anthony and M. Blaxter, unpublished). All sequences have been deposited in EMBL/GenBank/DDBJ. The perl program 'MOTU_define.pl' (R. Floyd and M. Blaxter, unpublished; based on CLOBB (Parkinson *et al.* 2002)) was used to allocate the resulting high-quality sequences to MOTU, based on pairwise identity scores and a user-defined cutoff.

The MOTU_define.pl program adds sequences one at a time to a growing database of barcode sequences (figure 1). It is a very simple procedure, internally consistent, and has the benefit of allocating stable MOTU identifiers to the dataset. As more sequences are generated, they can be added incrementally to the existing MOTU sets and thus continuity between experiments is attained. Indeed, sequence data can be acquired from other sources (such as GenBank/EMBL) and added to the dataset without compromising or changing the MOTU

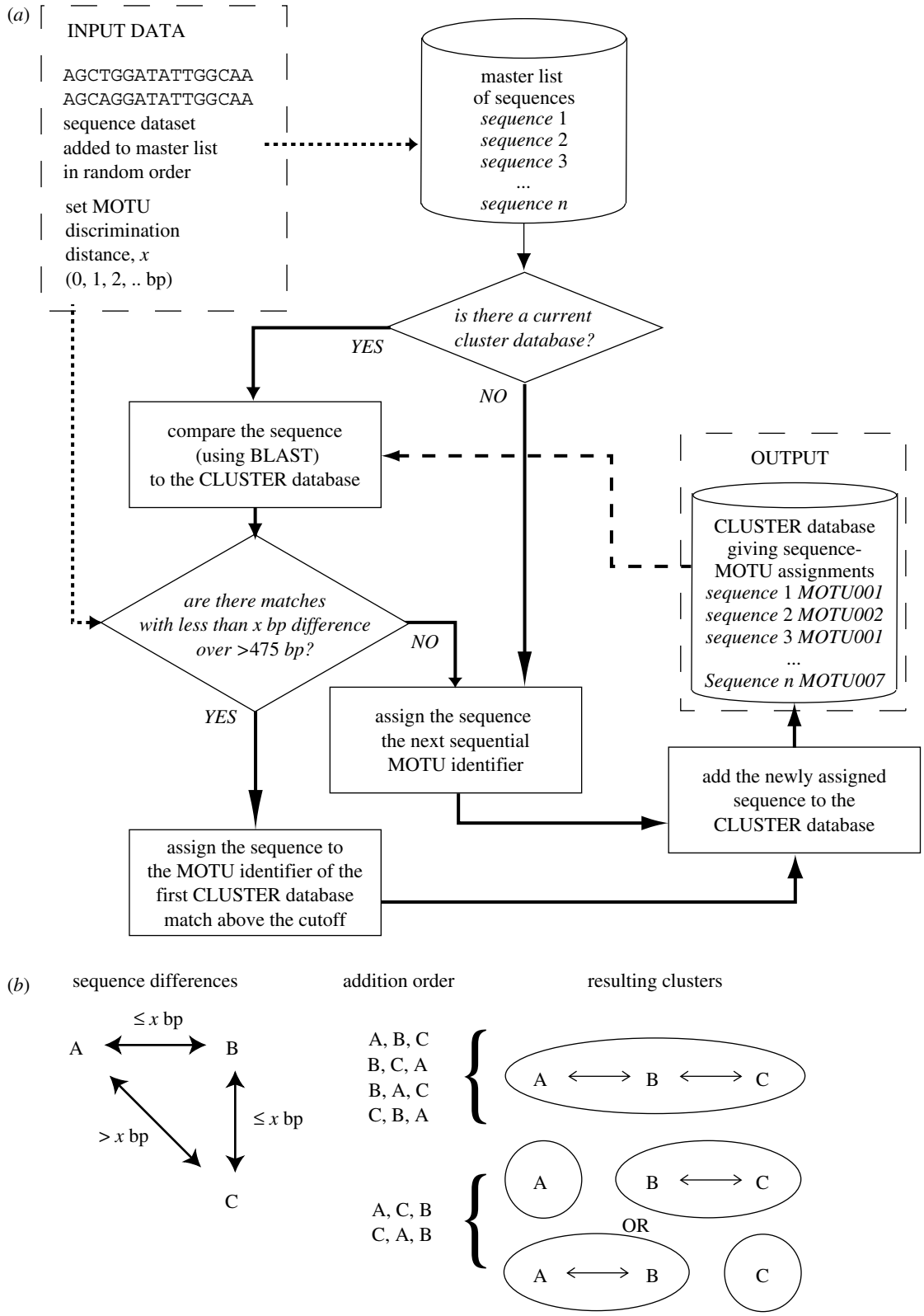


Figure 1. The MOTU_define.pl system. (*a*) A schematic of the process by which MOTU_define.pl allocates sequences to MOTU. The process can be run any number of times with different sequence addition order to assess MOTU stability. (*b*) The effect of addition order on MOTU definition. Three sequences, A, B and C, are clustered into MOTU. A differs from B, and B from C by less than the MOTU discriminant cutoff, but C differs from A by more than the cutoff. Depending on the order of analysis of the sequences, either one or two MOTU will be defined.

assignment of local data. This sort of process is ideal for building up a shared database of MOTU assignments and sequences. It is relatively rapid and reasonably scaleable (a variant of the program using the megaBLAST algorithm can cluster 100 000 expressed sequence tag sequences in ~20 h on a desktop computer (Parkinson *et al.* 2002); we do not yet have barcode datasets of this magnitude to test). The MOTU_define.pl is freely available from M. Blaxter, and requires only perl and a local copy of the NCBI BLAST suite (it is thus installable on UNIX, MacOSX and Windows systems).

Following our previous analyses of similar data, and our measured error rate in sequencing (~1 base in 3500) (Floyd *et al*. 2002), we standardly use a cutoff of 2 base differences in ~500 bp of sequence to discriminate MOTU: this can be varied. The program can also be rerun multiple times over the same set of sequences, randomizing the input order each time, and thus can be used to identify sequences and MOTU that do not behave simply under the cutoff statistic used. The use of single linkage clustering in MOTU_define.pl (where each sequence is clustered based on its identity to a single comparator) avoids issues of ambiguous alignment across a wide range of distantly related sequences. The high-quality sequences were aligned to each other and to a set of relevant control sequences from named taxa derived from GenBank or our previous studies and the alignment analysed using Maximum Parsimony in PAUP* 4.0b10 (Swofford 1999).

4. RESULTS: ETTRICK MOSS MEIOFAUNA

The moss fauna included animals from five animal phyla: Arthropoda (mites and collembolans), Tardigrada, Annelida (enchytraeids), Nematoda and Rotifera. The filtrates also included many protozoa (ciliates and amoebas) and some plant material. There was doubtless also a thriving unicellular fungal and algal, and prokaryote presence. Nematodes were most abundant, followed by rotifers and tardigrades (a ratio of 132 ± 20.8 nematodes to 6 ± 0.6 rotifers to 3 ± 0.9 tardigrades; mean and standard error of four samples corresponding to 0.5% of the extract from ~1 g dry weight of moss ecosystem); collembolans, mites and enchytraeids were rare in the moss, and excluded by their size from the filtrate.

(a) *Barcode sequence generation from single specimens*

Barcode sequences were derived from single specimens of nematodes, mites, collembolans, and enchytraeids. A total of 121 *cox1* sequences were generated from over 270 tardigrade specimens. For all taxa except rotifers, nSSU PCR and sequencing was successful ~85% of the time. In contrast, the *cox1* success rate was less than 40%. Indeed examination of available *cox1* sequences from animals related to those expected to be found in the moss ecosystem revealed that the 'universal' primers employed were unlikely to be able to amplify from some phyla. We conclude that use of the *cox1* target for the full diversity of animals will require additional rounds of primer pair optimization. No PCRs were successful from individually extracted rotifers, despite the nSSU primers sites being present in the available rotifer nSSU sequences. While this result could be due to the low number of cells (and thus genomes) in an individual rotifer, sequences from the bulk DNA sample were also rotifer-free (see below). We conclude that we will have to improve our extractions specifically to enhance rotifer DNA recovery.

(b) *Barcode sequences from nSSU libraries from bulk DNA*

A total of 145 sequences were generated from the bulk nSSU PCR library. Comparison to database sequences and single-specimen sequences from the same collection site (Blaxter *et al*. 2003) indicated that most derived from nematodes (123 or 85%) and four from tardigrades (3%). This ratio corresponds to that derived from the visual survey, excepting that no rotifer nSSU was recovered. In addition to these animal sequences, we isolated 18 nSSU sequences that clearly derived from ciliate protozoa, though none had an exact match in the public databases. We presume that these DNA segments were amplified because our primer set is not strictly metazoan-specific (we know that we can amplify environmental fungi, data not shown) and because, despite their being unicellular protozoa, ciliate macronuclei contain a many thousand fold amplification of the genes archived in the micronucleus, including the ribosomal RNA operons. No enchytraeid or arthropod sequences were recovered because the filtration excludes these larger meiofauna. Chimaeric amplicons are the bane of environmental sampling PCR. They arise from mispriming by amplification products during PCR, and result in DNA sequences that match one taxon at the 5′ end and another, unrelated one at the 3′ end. No chimaeric amplicons were identified, based on finding no discrepant BLAST matches for the first 250 compared to the last 250 bases of each.

(c) *Comparing single specimen and bulk nSSU MOTU*

MOTU_define.pl was used to infer MOTU from the nSSU datasets using a 2 bp difference cutoff. Data from the bulk sample and the single specimen sequences were clustered independently. For each nSSU MOTU, we derived a consensus sequence to represent that cluster for subsequent phylogenetic analysis (figure 2; but note that the definition of membership of a MOTU is not based on phylogenetic analysis). The use of a consensus sequence does not imply that this sequence correctly represents some ideal version of the true sequence, but rather is used to represent the diversity of the constituent sequences. The most abundant nSSU MOTU, derived from the bulk dataset, has 106 representatives, and is most similar to the chromadorid nematode *Plectus aquitilis*. Two of 16 single-specimen nSSU MOTU were also found in the bulk sample data (the *P. aquitilis*-like MOTU and a *Clarkus* (nematode)-like MOTU; figure 2). The bulk sequence dataset reflects the expected distribution of animals observed, excepting the Rotifera, and comparison with other more extensive datasets from soils and moss environments affirms that within the phyla that were amplified there is no apparent phylogenetic bias. Examination of this dataset suggests that the rate of identification of novel taxa using the barcode is not yet at saturation, despite the presence of the hyperabundant *P. aquitilis*-like Bulk_2bp_MOTU0001/Sin_2bp_ MOTU0005 (58% of all sequences, and 73% of the bulk sample sequences). Presumably, the rate of new MOTU identification could now be enhanced by prescreening for *P. aquitilis*-like sequences.

(d) *Comparison of cox1* and nSSU barcode analyses

A representative MOTU definition set for the *cox1* sequences is shown in figure 3. Twenty-two MOTU were defined, containing from 1 to 65 sequences

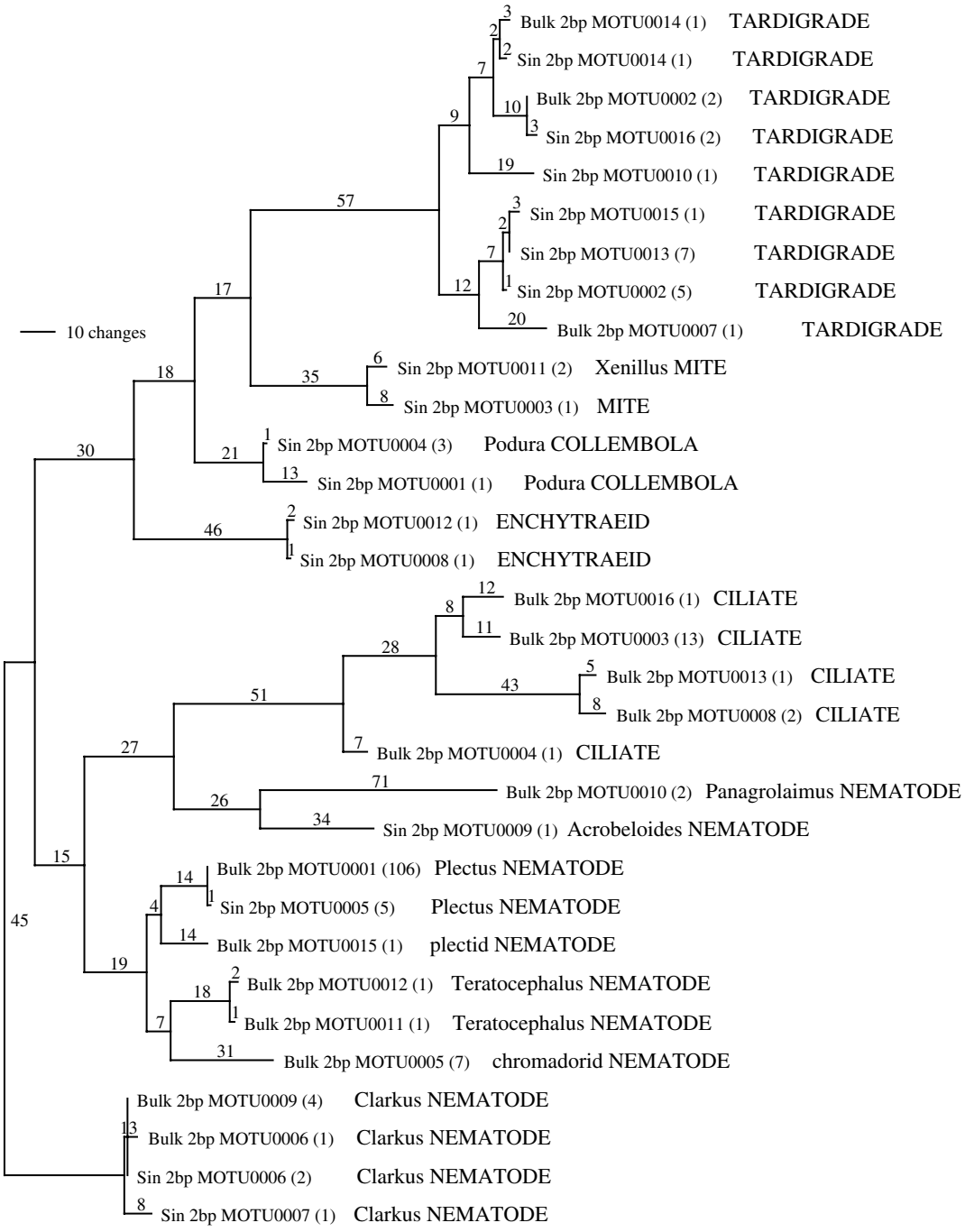


Figure 2. Meiofaunal MOTU defined using nuclear SSU sequences. The bulk nSSU dataset (Bulk) and a corresponding single specimen dataset (Sin) from the same moss sample were clustered into MOTU separately using a 2 bp cutoff, and consensus sequences predicted for those MOTU with more than one member. The consensus sequences and the singleton MOTU sequences were aligned and analysed using parsimony. For each MOTU represented the number of constituent sequences is given in brackets, and the taxonomic assignment based on BLAST search similarity to database sequences is given in bold. Where a taxon is identified below the major group, the MOTU sequence nested within a clade of sequences with the more specific designation (data not shown). Inferred numbers of changes are shown above each branch. Note that the tree is unrooted.

(figure 3*b*). The distribution of abundances of taxa implies one abundant taxon (~50% of the sample) and a larger number of taxa with low abundance.

For 82 tardigrade specimens, we obtained sequences of both *cox1* and nSSU with >490 bp of high-quality data. The two markers were used to infer independent clusterings, using a 2 bp cutoff, and the resultant clusters compared (figure 4). Seventeen *cox1* MOTU were defined from this subset. Surprisingly, 23 nSSU MOTU were defined, despite the overall lower level of sequence divergence, though the distance between distinct clusters was greater in the *cox1* dataset (as would be expected from the known higher substitution rate in animal mitochondrial genes). Seven MOTU with single members were found in both datasets, and two *cox1* MOTU (with two and five members) corresponded to two nSSU MOTU each (figure 4). The remaining 68 specimens formed two groups with complex patterns of overlap between nSSU and *cox1* MOTU (figure 4). Thus, while *cox1*

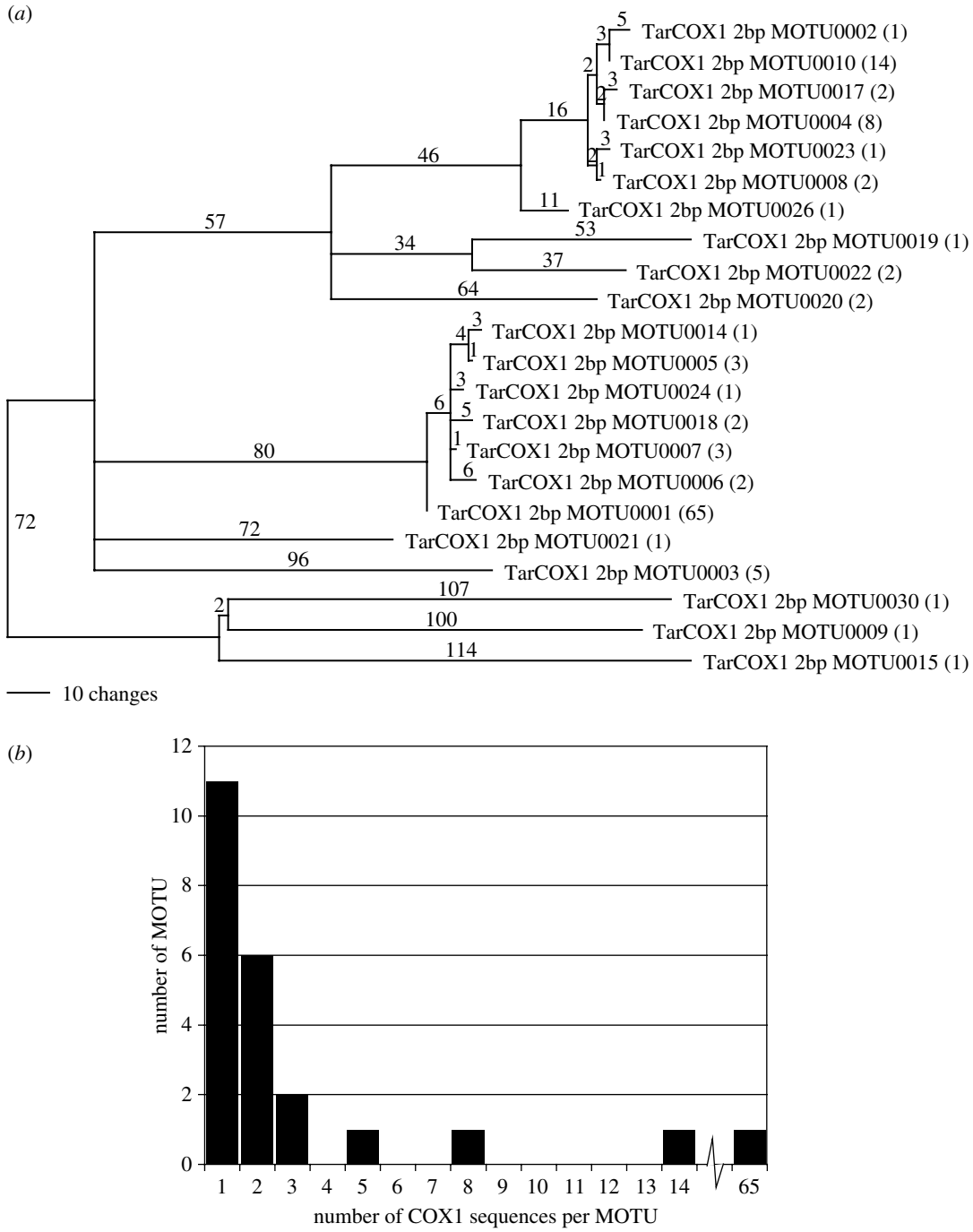


Figure 3. Tardigrade MOTU defined using *cox1* sequences. (*a*) A consensus sequence was derived for each MOTU, and these were aligned. The branch lengths are proportional to the number of discrete changes mapped to each. The number of sequences assigned to each MOTU is given in brackets after the MOTU name. (*b*) Histogram of MOTU abundance in the 121-sequence *cox1* dataset.

and nSSU are both effective at defining MOTU, and there was a general agreement between the two cluster sets, there were also significant disagreements. Whether these disagreements are due to the population history and hybridization patterns of the specimens sampled or are indicators of real incongruence between the markers is not clear. The two clouds of taxa (marked in figure 4) may correspond to particularly variable single taxa, or perhaps diverging radiations of taxa. Many tardigrades can reproduce asexually, or have sex only very rarely (Kinchin 1994), and thus this pattern may reflect divergence of clonal or matrilineal lines.

5. RESULTS: PROPERTIES OF EXACT SCORE MOTU DEFINITION

(a) *Variability due to single linkage clustering*

Assignment of any single sequence to a MOTU depends critically on what sequences have been added previously (figure 1*b*). If one takes three sequences, where only two differ by more than the chosen cutoff, the order of addition changes the number and membership of MOTU inferred. Rather than being a failing of this procedure, we regard this as being a feature: it permits exploration of the 'clouds' of taxa that are closely related. If a set of specimens robustly clusters into a particular set of MOTU, no

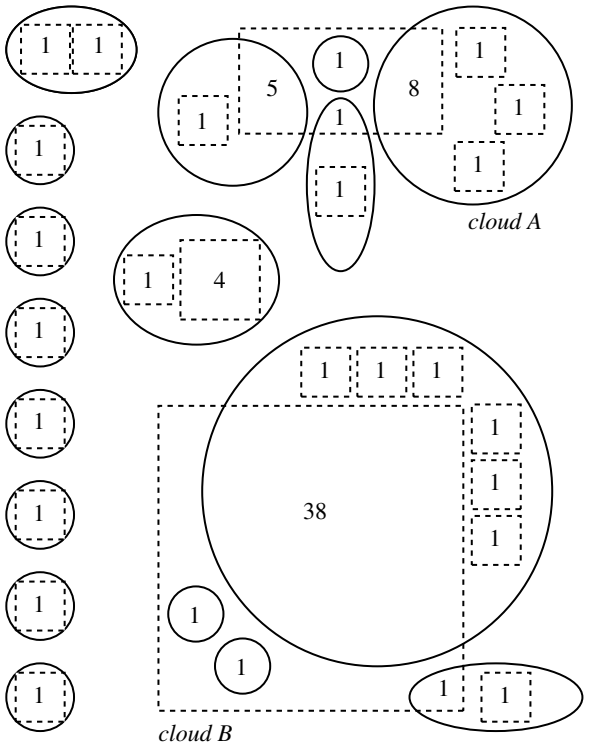


Figure 4. Comparison of MOTU definition using nSSU and *cox1* markers. This Venn diagram shows *cox1* MOTU sets (solid circles) and nSSU MOTU sets (dotted squares). The numbers within each partition indicate the number of individual specimens (out of 82) placed there.

matter what the addition order, this suggests that these MOTU have some congruence with biological taxa, and a distinctness from other related OTU. But if repeated clustering of a group of sequences yields discordant MOTU, this identifies a biologically interesting phenomenon, somewhere along the spectrum from population genetic processes to recently separated taxa still sharing ancestral polymorphisms. Such variability can thus alert researchers to novel features of communities not simply accessible through other means.

(b) *MOTU inference using different cutoff scores*

We performed 300 independent clusterings of the 295 tardigrade nSSU sequences. One hundred independent, random-addition order replicates were produced for taxon definition cutoffs of 2, 3 and 4 bp. For the 2 bp cutoff, the number of MOTU inferred ranged from 143 to 157, with a mode of 151 and a mean of 149.96 ± 2.61. The majority of the variability in MOTU number inferred was due to alternate groupings of a few clouds of sequences (not shown). The use of larger cutoff values also resulted in MOTU sets with wide ranges ($\sim$10% of the total number inferred) (figure 5). Thus increasing the fuzziness of the MOTU discriminant does not result in a simple collapse of the clouds of sequences into single taxa. We have also observed this pattern in other meiofaunal datasets (Blaxter *et al.* 2003) (Floyd, Blaxter *et al.* unpublished).

The variability of attribution observed between independent clusterings is not a unique feature of MOTU_define.pl: the same issue must arise in all other methods, as the data we are using to infer taxa is essentially quantal.

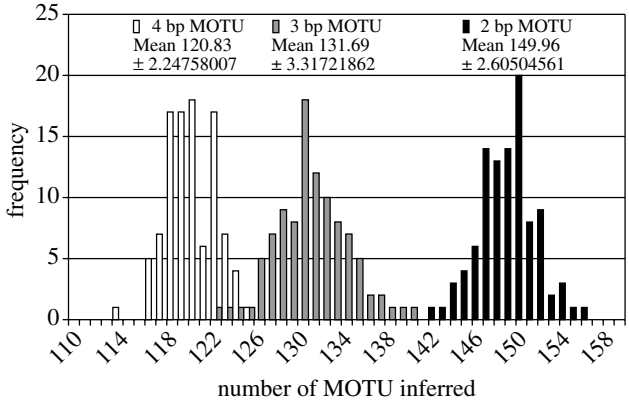


Figure 5. Variability in the number of MOTU defined by replicate analyses. The histogram shows the frequency distribution of total numbers of clusters inferred from 295 tardigrade nSSU sequences from the Glen Ettrick study site using MOTU_define.pl at three different cut off values: 2 bp (black), 3 bp (hatched) and 4 bp (open). The mean and standard deviation of each set of analyses is given.

6. DISCUSSION: TURNING SEQUENCES INTO MOTU

The MOTU-define.pl program is but one, obvious way of inferring MOTU. Other methods could also be applied. One common choice is to use a BLAST algorithm (usually BLASTn) (Altschul *et al.* 1997) to identify the best match in a reference database, and to assign the identifier of the best match to the barcoded specimen if the similarity is judged to be good enough. This method has many pitfalls, not least its reliance on a well-populated (and correctly named) database of barcodes. In meiofaunal surveys such as presented here, the lack of close relatives in the database can make this approach less-than-rewarding. More importantly, the BLAST algorithm (Altschul *et al.* 1990; Altschul *et al.* 1997) was not designed for barcode identity assignment, and simply taking the top-scoring match, with some predefined quality score cutoff, may miss issues of, for example, equal top scoring matches. A variation on the BLAST approach would be to extract the best matches (for example, all matches with a score within a small percentage of the best match), perform a complete alignment with the barcode query sequence, and then subject this alignment to model-driven phylogenetic analysis to ask if the barcoded specimen is a credible member of a monophyletic clade with any of the references.

Because much DNA barcode sequence is derived from single sequencing reads on only one strand of the DNA, the quality of the sequences may not be as good as those in the databases. The sequencing chromatogram can be analysed to yield a quality score for each base (Ewing & Green 1998; Ewing *et al.* 1998), and these could be incorporated into a BLAST-and-align method for MOTU definition that down-weights any differences associated with low quality scores and pays more attention to high-quality scores. A variation on this method might also include partitioning the aligned

sequences *a priori* into more- and less-informative sites. Thus, in a protein-coding gene such as *cox1*, one might give first and second base changes more weight than those observed in fourfold degenerate sites. In a RNA gene such as nSSU, one could differentially weight residues by their involvement in secondary structure, and their observed conservation in large aligned datasets.

As barcoding is applied somewhere on the span between population genetics and taxon phylogenetics, the use of network-based algorithms may also assist. Templeton network analysis is much used in population studies to examine patterns of haplotype distribution and relatedness (Clement *et al*. 2000). For DNA barcode data, such network analysis, with different cutoffs for the breaking of ties between subnetworks, can assist in understanding the patterns of diversity in the sequences and thus the likely status of the MOTU defined. In genomics, definition of protein families has been achieved using multiple cluster linkage methods, where complex networks of similarity between sequences can be examined at different levels of granularity to identify coherent clusters (Enright *et al*. 2002). A similar approach applied to DNA barcode data might be doubly informative of not only final MOTU but also the interrelationships of MOTU clouds.

Ultimately, we might want to use rigorous phylogenetic methods to affirm the monophyly of our newly defined MOTU, and to place them in the context of named sequence diversity. However, we must be aware of the issues of partial sorting of haplotypes between lineages as they diverge. Wide-ranging studies on several taxa have clearly shown that while rapidly evolving sequences are very well suited to generation and testing of taxon hypotheses at local scales, they are often very much unsuited to deeper phylogenetic analysis. Processes such as base substitution bias and variable site saturation can rapidly obscure real phylogenetic signals and generate spurious trends in data. The barcode data will be rather unsuitable for reconstructing the deeper branches of the tree of life, including perhaps all those below the generic level (Vogler *et al*. 2005). Simply using trees to infer taxa from barcode data can be positively misleading: we should rather define the taxa and then examine their relationships through rigorous phylogenetics.

Taxa defined by MOTU methods can be used for standard taxonomic and ecological surveys. By comparing the barcode sequence with a database of sequences from specimens identified to Linnaean taxa before sequencing, the anonymous survey specimens can be placed within the known taxonomic framework, and the organismal biology of the organisms from which they derived inferred (Floyd *et al*. 2002; Blaxter & Floyd 2003; Blaxter 2004). By this method we can move from anonymous sequence to ecosystem biology.

This work was carried out as part of ongoing investigations into meiofaunal diversity in our laboratory, and was funded by the UK Natural Environment Research Council and the Linnaean Society of London. J.M. and T.C. carried out the meiofaunal surveys as part of their major undergraduate projects.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic local alignment search tool. *J. Mol. Evol.* **215**, 403–410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402. (doi:10.1093/nar/25.17.3389.)

Amaral Zettler, L. A., Gomez, F., Zettler, E., Keenan, B. G., Amils, R. & Sogin, M. L. 2002 Microbiology: eukaryotic diversity in Spain's River of Fire. *Nature* **417**, 137. (doi:10.1038/417137a.)

Berry, W. 1970. *Farming: a handbook*. San Diego, CA: Harcourt Brace & Company.

Blaxter, M. 2003 Molecular systematics: counting angels with DNA. *Nature* **421**, 122–124. (doi:10.1038/421122a.)

Blaxter, M. L. 2004 The promise of a molecular taxonomy. *Phil. Trans. R. Soc. B* **359**, 669–679. (doi:10.1098/rstb.2003.1447.)

Blaxter, M. L. & Floyd, R. 2003 Molecular taxonomics for biodiversity surveys: already a reality. *Trends Ecol. Evol.* **18**, 268–269. (doi:10.1016/S0169-5347(03)00102-2.)

Blaxter, M., Elsworth, B. & Daub, J. 2003 DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Biol. Lett.* **271**, S189–S192. (doi:10.1098/rsbl.2003.0130.)

Blaxter, M., Floyd, R., Dorris, M., Eyualem, A. & De Ley, P. 2004 Utilising the new nematode phylogeny for studies of parasitism and diversity. In *Nematology monographs and perspectives* (ed. R. Cook & D. J. Hunt), pp. 615–632. Leiden: E. J. Brill.

Blaxter, M. L. *et al*. 1998 A molecular evolutionary framework for the phylum Nematoda. *Nature* **392**, 71–75. (doi:10.1038/32160.)

Clement, M., Posada, D. & Crandall, K. A. 2000 TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**, 1657–1659. (doi:10.1046/j.1365-294x.2000.01020.x.)

De Ley, P. & Bert, W. 2001 Video capture and editing as a tool for storage, distribution and illustration of morphological characters of nematodes. *J. Nematol.* **34**, 296–302.

De Ley, P. *et al*. 2005 An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Phil. Trans. R. Soc. B* **360**, 1945–1958. (doi:10.1098/rstb.2005.1726.)

Diez, B., Pedros-Alio, C. & Massana, R. 2001 Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* **67**, 2932–2941. (doi:10.1128/AEM.67.7.2932-2941.2001.)

Enright, A. J., Van Dongen, S. & Ouzounis, C. A. 2002 An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584. (doi:10.1093/nar/30.7.1575.)

Ewing, B. & Green, P. 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

Ewing, B., Hillier, L., Wendl, M. C. & Green, P. 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.

Eyualem, A. & Blaxter, M. 2003 Comparison of biological, molecular and morphological methods of species identification in a set of cultured *Panagrolaimus* isolates. *J. Nematol.* **35**, 119–128.

Finlay, B. J. 2002 Global dispersal of free-living microbial eukaryote species. *Science* **296**, 1061–1063. (doi:10.1126/science.1070710.)

Floyd, R., Eyualem, A., Papert, A. & Blaxter, M. 2002 Molecular barcodes for soil nematode identification.

Mol. Ecol. **11**, 839–850. (doi:10.1046/j.1365-294X.2002.01485.x.)

Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. 1990 Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63. (doi:10.1038/345060a0.)

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**, 313–321. (doi:10.1098/rspb.2002.2218.)

Kinchin, I. M. 1994. *The biology of tardigrades*. London: Portland Press.

Lambshead, P. J. D. 1993 Recent developments in marine benthic biodiversity research. *Oceanis* **19**, 5–24.

Lambshead, P. J. D. & Boucher, G. 2003 Marine nematode deep-sea biodiversity—hyperdiverse or hype? *J. Biogeo.* **30**, 475–485.

Lawton, J. H. *et al.* 1998 Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forests. *Nature* **391**, 72–75. (doi:10.1038/34166.)

Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C. & Moreira, D. 2001 Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607.

Maucci, W. (ed.) 1986 *Fauna d'Italia Tardigrada*. Bologna: Edizioni Calderini Bologna.

May, R. M. 1988 How many species are there on earth? *Science* **241**, 1441–1449.

Moreira, D. & Lopez-Garcia, P. 2002 The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends Microbiol.* **10**, 31–38. (doi:10.1016/S0966-842X(01)02257-0.)

Parkinson, J., Guiliano, D. & Blaxter, M. 2002 Making sense of EST sequences by CLOBBing them. *BMC Bioinformatics* **3**, 31. (doi:10.1186/1471-2105-3-31.)

Platt, H. M. 1994 Foreword. In *The phylogenetic systematics of free-living nematodes* (ed. S. Lorenzen). London: The Ray Society.

Swofford, D. 1999. *PAUP** 4.b10. Sunderland, MA, USA: Sinauer Associates.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. 2003 A plea for DNA taxonomy. *Trends Ecol. Evol.* **18**, 70–74. (doi:10.1016/S0169-5347(02)00041-1.)

Vogler, A., Cardoso, A. & Barraclough, T. 2005 Exploring rate variation among and within sites in a densely sampled tree: species level phylogenetics of north american tiger beetles (genus *Cicindela*). *Syst. Biol.* **54**, 4–20.

Chapter 3

Nematode DNA Barcoding: Comparisons of the performance of different markers

3.1 Introduction

Molecular diagnosis of taxonomic affinities of specimens is now a common practice across biological research disciplines. The idea of using molecular barcodes (Floyd *et al.*, 2002; Hebert *et al.*, 2003a, 2003b) for designation of taxa is now widely accepted although the definition of a taxon (or what constitutes a species) is still a matter of discussion. It is now common practice with high-throughput projects routinely producing hundreds if not thousands of barcodes. Initiatives such as the Consortium for the Barcoding of Life (CBOL, http://www.barcoding.si.edu/) produce much data and have also generated numbers of daughter projects such as FISH-BOL (http://www.fishbol.org/) and the All Birds Barcoding Initiative (ABBI, http://www.barcodingbirds.org/). Barcodes have been used for various purposes from forensic identifications (Lorenz *et al.*, 2005) to large-scale environmental diversity surveys (Venter *et al.*, 2004; Sogin *et al.*, 2006).

Broadly speaking, two current uses exist for the application of barcodes. Firstly, they can be used for confirmation. Whether regulatory, forensic or investigatory, a barcode is associated with a particular taxonomic scheme. So a novel sequence can be assigned by asking whether it is the same (or within certain limits can be considered the same) to a barcode previously generated. Secondly, barcodes can be inferential. They can be used in specimen-independent environmental surveys to assess the molecular diversity of organisms present (Blaxter *et al.*, 2005). Further questions can be asked of these well-differentiated groups, for example, are they of any biological significance i.e. do they represent putative species, or are they populations within species?

Failure to delineate species can be a result of a lack of variation among their barcode sequences. On occasion, this can be overcome by more intensive sampling, by increasing the number of individuals sampled from a taxon to

reveal the intraspecific variation. Given intraspecific variation, it is important to generate sequence from many specimens per taxon so divergent specimens can be correctly allocated.

However, there are situations where increased sampling will not aid resolution (e.g. (Trewick), 2008). Geographically and recently speciated taxa may not have accumulated sufficient differences to allow resolution from barcodes (Elias *et al.*, 2007; Wiemers and Fiedler, 2007) and may retain ancestral polymorphism. This situation may be compounded if the species-definition used is incorrect (e.g. morphological traits which are unable to differentiate cryptic species). In this case, increased sampling will still result in unresolved taxa as the molecular taxa will not be congruent with the morphological taxa.

Meiofauna, organisms with a maximal body axis of less than 1 mm, are hyper-abundant and include nematodes, tardigrades and rotifers (Lawton *et al.*, 1998). Traditional morphological species definitions are difficult to apply to meiofauna due to their microscopic size. Further, they are ubiquitously distributed and some taxa exhibit high levels of molecular diversity (Blaxter *et al.*, 2004). These factors will challenge the efficiency of molecular barcoding as a tool for meiofaunal identification.

Consideration must be given to the gene selected for barcoding. There need to be conserved regions so that universal primers can be used, and variable regions are also required to give taxonomic information.

Placing all inferences solely on the results from one gene is unlikely to reflect much more than the gene history. Results from some recent studies proposed that at least two genes should be used, preferably independent of each other (i.e. one nuclear and one mitochondrial) (Elias *et al.*, 2007).

Hebert *et al.* (2003a; 2003b) have proposed the mitochondrial marker cytochrome oxidase I (CO1) and this has been adopted by CBOL as the barcoding standard. However its 'universality' is under scrutiny (Lorenz *et al.*, 2005) and it has been suggested that is unreliable for inferring phylogenetic relationships (Moritz and Cicero, 2004; Hurst and Jiggins, 2005; Sonnenberg *et al.*, 2007). As the third base position is less constrained than

the first and second, variable site saturation can obscure phylogenetic signal. Whilst deep phylogenetic resolution may not be necessary for barcoding, using barcodes to generate trees, which are then used to define species, should be approached with caution. Likewise, relying solely on a nuclear marker (such as large or small ribosomal subunit LSU and SSU respectively) may not provide enough resolution between taxa.

When COI was proposed as a standard universal marker, previous work (Johns and Avise, 1998) suggested that a 2% difference was sufficient to discriminate closely related vertebrate species based on results from mitochondrial cytochrome b (*cytb*). Hebert *et al*. (2003b) investigated the COI divergences across species pairs representing 11 animal phyla and found divergences for congeneric species pairs up to 53.7% for COI. Whilst most taxa showed interspecific differences of more than 8% (Hebert *et al.*, 2003b) there were some taxa that exhibited much lower COI differences (e.g. cnidarians showed less than 2% difference between species (Hebert *et al.*, 2003b)); and others that have much greater differences (e.g. amphibians; 10-14% for mantellid frogs (Vences *et al.*, 2005)).

The idea of a barcoding gap underpins all molecular barcoding studies. Molecular barcodes rely on differences or similarities of DNA sequences to separate or cluster sequences into taxa. Small differences (less than 1%) are assumed to represent individual differences (intraspecific variation), while larger differences (approximately 2%) are assumed to reflect distinctions between different species (interspecific divergence). In order for barcoding to work, there should be no overlap between the intraspecific variation and interspecific divergence (Meyer and Paulay, 2005). So when two sequences are compared, the amount of genetic distance between them would indicate if they belonged to the same species. For COI data from cowries (Meyer and Paulay, 2005), there was an overlap between intra- and interspecific variation and a threshold could not easily be chosen to define taxa without the risk of over-splitting or lumping taxa. Other taxa also fail to show a barcoding gap (Elias *et al.*, 2007; Wiemers and Fiedler, 2007).

Once a dataset has been generated (preferably using multiple nuclear and mitochondrial genes for a complementary dataset), there is no one standard

way of assigning sequences to taxa. There is a suite of different analytical methods to develop taxa from. One is to classify sequences into taxa by inferring a phylogenetic tree and choosing taxa using various parameters. Ultimately clades are defined by visual inspection of trees by the investigator. This is a highly subjective method as one person's well-defined clade is another's nested subpopulation.

An objective method would designate taxa purely on the information contained within the sequence data. One method used is the basic local alignment search tool (BLAST) to search for most similar named sequences. This method is dependent upon a broad ranging and correctly identified database from which to search from (e.g. GenBank). An issue remains regarding the cut-off to use. Alternatively, sequences can be algorithmically clustered into operational taxonomic units (OTUs) independent of any species designation.

In this study we investigated the behaviour of one mitochondrial and two nuclear genes (mtCO1, nLSU and nSSU) for definition of molecular operational taxonomic units (MOTUs) in a sample of terrestrial nematodes. We used these genes to assess the use of two algorithms for clustering the sequence data, MOTU_define.pl and DOTUR. The data were also interrogated for the presence of a barcoding gap.

3.2 Methods

3.2.1 Sample Collection

The samples investigated were collected by Dr Asher Cutter. Nematodes were extracted from soil, vegetation and invertebrate samples collected from various locations (Table 3.1). Cultures were founded by placing single females on NGM agar plates, seeded with *E. coli* OP50. For successful cultures, individual offspring were picked for lysis using a direct NaOH procedure (Floyd *et al.*, 2002) in a final volume of 40 μl. For those that failed to reproduce, the single adult was picked for lysis.

3.2.2 PCR and Sequencing

Three genes were chosen as PCR targets: the 5′ end of mitochondrial cytochrome oxidase I (COI), the 5′ D2-D3 loop region of the nuclear ribosomal large subunit (LSU) gene and the 5′ end of the nuclear ribosomal small subunit (SSU) gene. Primer sequences, expected length of product and PCR cycles are listed in table 3.2. For all amplifications, 2 μl of lysate was used in a 20 μl reaction with 2 μl 10x PCR buffer, 2 μl 0.2 mM dNTPs, 0.4 μl 10 μM of each primer and 0.08 μl Taq polymerase (Qiagen).

All PCR products were visualized by separation on 1.5% agarose gels stained with 0.0002% ethidium bromide with 1 kb DNA ladder (Invitrogen Corp., Carlsbad, CA, 92008, USA) as size markers. Positive products were cleaned using Exonuclease I (New England Biolabs, Inc., Ipswich, MA 01938, USA) and Shrimp Alkaline Phosphatase (USB Corp., Cleveland, OH, 44128, USA) following the Wellcome Trust Sanger Institute protocol and sequenced using the primers LCO1490 or C1-J-1718, D2A and SSU_R_09.

Table 3.1 Summary of sample data. Alternative names represent samples that had been used to generated sequences before this study. Accession numbers are given for sequences.

Location	ID	Alternative name	COI	LSU
Chile	01	SA-TF1-A-07	-	FN386615
	02	SA-TF1-A-06	FN397751	FN386616
	03	SA-TF1-A-05	FN397752	FN386617
	04	SA-TF1-A-04	FN397753	FN386618
	05	SA-TF1-A-03	FN397754	FN386619
	06	SA-TF1-A-02	FN397755	FN386620
	07	SA-01-B-20	-	FN386621
	08	SA-01-B-17	FN397756	FN386622
	11	SA-WP12-C-02	FN397757	FN386623
	12	SA-WP12-C-01	-	FN386624
	13	SA-01-A-06	FN397758	FN386625
	14	SA-01-A-05	FN397759	FN386626
	15	SA-01-A-03	FN397760	-
	16	SA-01-A-02	FN397761	FN386627
Taipei	17	4-a-1	-	FN386628
	18	3-a-7	FN397762	FN386629
	19	3-a-6	FN397763	FN386630
	20	3-a-5	-	FN386631
	21	3-a-3	FN397764	FN386632
	22	3-a-1	-	FN386633
	23	4-a-3	FN397765	FN386634
	24	3-a-2	FN397766	FN386635
	25	2-a-1	-	FN386636
	26	5-a-1	FN397767	FN386637
	27	7-a-1	FN397768	FN386638
	28	7-a-2	-	FN386639
	29	8-a-1	FN397769	FN386640
	30	8-a-2	FN397770	FN386641
	31	8-a-3	FN397771	FN386642
	32	8-a-4	FN397772	FN386643
	33	13-a-1	FN397773	FN386644
	34	13-a-2	FN397774	FN386645
	35	15-a-1	FN397775	FN386646
	36	15-a-2	-	FN386647
	37	15-a-3	-	FN386648
	38	15-a-4	FN397776	FN386649
	39	15-a-5	-	FN386650
	40	23-a-1	FN397777	FN386651
	41	4-a-4	FN397778	FN386652
	42	23-a-2	-	FN386653
	43	23-a-3	FN397779	FN386654
	44	23-a-6	FN397780	FN386655
	45	25-a-2	-	FN386656
	46	25-a-4	FN397781	FN386657
	47	25-a-5	FN397782	FN386658
	48	25-a-6	FN397783	FN386659
	49	11-a-1	-	FN386660
	50	25-a-3	FN397784	FN386661
	51	23-a-5	FN397785	FN386662
	52	23-a-4	FN397786	FN386663

	53	4-a-5	FN397787	FN386664
	54	4-a-2	FN397788	FN386665
Edinburgh	57	-	FN397789	FN386666
	58	-	FN397790	FN386667
	59	-	FN397791	FN386668
	60	-	-	FN386669
	61	-	-	FN386670
Taipei	62	25-a-1	-	FN386671
	63	13-a-3	-	FN386672
	64	11-a-2	FN397792	FN386673
Edinburgh	65	-	-	-
	66	-	-	-
	67	-	-	FN386674
	68	-	-	-
	69	-	-	-
	70	-	FN397793	FN386675
	71	-	FN397794	-
	72	-	FN397795	FN386676
	73	-	-	FN386677
	74	-	FN397796	FN386678
	75	-	-	FN386679
	76	-	-	FN386680
	77	-	-	-
	78	-	-	FN386681
	79	-	-	-
	80	-	FN397797	FN386682
	81	-	FN397798	FN386683
	82	-	-	FN386684
	83	-	FN397799	FN386685
	84	-	FN397800	FN386686
	85	-	-	FN386687
	86	-	-	FN386688
	87	-	-	FN386689
	88	-	-	-
	89	-	-	FN386690
	90	-	-	FN386691
	91	-	FN397801	FN386692
	92	-	FN397802	FN386693
	93	-	FN397803	FN386694
	94	-	-	FN386695
	95	-	-	FN386696
	96	-	-	-

Table 3.2 PCR primers and conditions used

Target gene	Primer name	Sequence (5' - 3')	Expected product length (bp)	PCR conditions	Reference
COI	LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G	658	94°C 60 sec; 5 cycles of 94˚C 60 sec, 45°C 90 sec, 72°C 90 sec; 35 cycles of 94˚C 60 sec, 50°C 90 sec, 72°C 60 sec; 72°C 300 sec	(Hebert *et al.*)
	HCO2198	TAA ACT TCA GGG TGA CCA AAA AAT CA			
	C1-J-1718	GGT GGT TTT GGT AAC TGA TTT TAC C	525	94°C 120 sec; 35 cycles of 94˚C 70 sec, 45°C 70 sec, 72°C 90 sec; 72°C 420 sec	(Read *et al.*)
	C1-J-2191	GCT GGT AAA ATC AAA ATA TAT ACT TC			
LSU	D2A	ACA AGT ACC GTG AGG GAA AGT TG	≈890	95°C 300 sec; 5 cycles of 94˚C 30 sec, 55°C 45 sec, 72°C 120 sec; 72°C 600 sec	(Ye *et al.*)
	D3B	TGC GAA GGA ACC AGC TAC TA			
SSU	SSU_F_04	GCT TGT CTC AAA GAT TAA GCC	893	94°C 300 sec; 35 cycles of 94˚C 60 sec, 55°C 90 sec, 72°C 120 sec; 72°C 10 sec.	(Blaxter *et al.*)
	SSU_R_26	CAT TCT TGG CAA ATG CTT TCG			
	SSU_F_07	AAA GAT TAA GCC ATG CAT G	888		Unpublished
	SSU_R_26				
	SSU_F_07		545		
	SSU_R_09	AGC TGG AAT TAC CGC GGC TG			(Blaxter *et al.*)

Sequences were collected on a 48-capillary ABI 3730 DNA Analyzer and sequence chromatograms were processed using trace2seq.pl (A. Anthony and M. Blaxter, unpublished). trace2seq uses phred (Phil Green, unpublished) to identify high-quality base calls and outputs sequence trimmed of low quality calls). Sequences were compared with published sequences in GenBank by means of BLAST search (Altschul *et al.*, 1997). Sequences were used in analysis if top BLAST hits were nematode sequences.

3.2.3 MOTU_define.pl

MOTU_define.pl (Floyd *et al.*, 2002) is a perl script that uses a rule-based discriminant to assign sequences to MOTU. The analyses performed here used MOTU_define.pl version 2.08 (M. Blaxter, J. Mann, R. Floyd, unpublished; see http://www.nematodes.org/bioinformatics/ for download), which permits re-sampling at multiple cut-offs. The script uses BLAST to identify the best matching, previously analysed sequence, and then uses a count of nucleotide differences to decide whether the new sequence should be assigned to a new MOTU or join an existing one. MOTU defined using MOTU_define.pl are permanent and can be added to incrementally, preserving previously assigned identifiers.

MOTU_define.pl was used to analyse each gene set separately, with a minimum sequence length of 230 bp (the smallest sequence obtained was LSU_51, which was 238 bases), a minimum overlap length of 150 bp (65% of min sequence length), 100 re-samplings and the range of cut-off values 0-10, 15, 20, 30, 40, 50, 55, 60 as the parameters for analysis.

3.2.4 DOTUR

DOTUR is an OTU definition program that assigns sequences to taxa based on an alignment and derived distance matrix (Schloss and Handelsman,

2005). We term the OTU derived from DOTUR as 'DOTU'. For each congruent gene set an alignment was generated using ClustalX. Distance matrices were generated using the Kimura 2-parameter (K2P) and Jukes-Cantor (JC) distance settings using PHYLIP version 3.67 (http://evolution.gs.washington.edu/phylip.html). Each gene dataset was processed using 1000 iterations, furthest and nearest neighbour clustering methods with a precision setting of 1000 (Schloss and Handelsman, 2005). Furthest neighbour (FN) clustering will only add a sequence to a DOTU if it is sufficiently similar to all other sequences in the DOTU. This is likely to generate more DOTUs than nearest neighbour (NN) clustering which will add a sequence if it is similar to any sequence in the DOTU.

3.2.5 Complete data set analysis

PCR success can be stochastic and not all specimens would be expected to yield sequences for all three genes. Specimens that would, would allow direct comparison between the genes. However, analysis of the entire gene sets would highlight any clustering that may be caused by only using a subset of the data. Additional sequences may result in different OTUs forming (either splits, joins or complex re-distributions of sequences. Conversely, equivalent MOTUs may be generated with the extra sequences joining established MOTUs. For each gene set, a multiple alignment was constructed using ClustalX (with default settings). The alignments were used to generate neighbour joining (NJ) trees in PAUP using absolute differences as distance settings.

3.3 Results

3.3.1 PCR and Sequencing

From the collection of 92 lysates, 53 COI, 82 LSU and 73 SSU nematode sequences were generated before lysates were exhausted (57.6%, 89.1% and 79.3% success respectively). Sequences were only used if the top BLAST hits were nematode sequences. Most SSU sequences (52) had been previously generated by Dr Cutter before the samples were used in this study. These were compared with newly generated sequences (21) to ensure that specimen-order followed the order used by Dr Cutter. The COI primer pair LCO1490 and HCO219 was initially used for PCR, producing 43 sequences. Samples that did not amplify were then tried with C1-J-1718 and C1-J-2191 (generating 10 positive results) however there were 35 samples that failed to amplify with either COI primer pair (other positive PCRs generated non-nematode sequences). Despite the support for COI as the universal barcode (Hebert *et al.*, 2003a), these results suggest that COI is not always as readily available for barcoding as previously thought. SSU primer pairs also had variable success, the most successful primer pair was SSU_F_07 and SSU_R_09, generating 69 out of 72 sequences in the complete data set.

The three gene sets were filtered to generate a congruent data set where sequences for all three genes had been recovered for the same sample. This consisted of 48 sequences for each gene. The mean sequence length and standard deviation for congruent sequences for COI, LSU and SSU was 533.9 (± 95.7) bp, 558.7 (± 68.1) bp and 430.3 (± 30.2) bp respectively.

3.3.2 MOTU results

MOTU_define.pl was used to analyse each congruent gene set separately. The number of MOTUs and clustering behaviour was examined over the range of cut-offs used (Figure 3.1).

For COI, the number of MOTUs defined decreased as the cut-off (number of base pairs) was increased. The mean number of unique sequences, i.e. the number of MOTUs using 0 bp as a cut-off value, was 44.06 ± 0.5. Initially there was a sharp decrease in the number of MOTU defined as the cut-off value was increased to 3 bp (Figure 3.1). The rate of MOTU definition then decreased as the cut-off value approached 10bp. There was then little change in the number of MOTU defined as the line plateaus. Between 15 and 20 bp cut-offs, the mean number of MOTUs defined was stable at 13 MOTUs (15 bp cut-off defined 13.24 ±0.48, 20 bp cut-off defined 13.00 ±0.48) and 25 bp cut-off defined 12.62 ±0.73 MOTUs. After 30 bp there was another drop in the number of MOTU defined as the threshold allowed less similar sequences to be assigned to the same MOTU. At a cut-off of 60 bp (11.2% of mean sequence length), the number of MOTU defined was 5.75±1.03.

At the plateau phase the number of MOTU defined were stable. If this is to be used as an indicator of a barcoding gap, then the members of MOTUs should also be stable. The members of MOTUs were investigated for all re-samples at 15 bp cut-off (Figure 3.2).

From the re-sample data, eleven MOTUs were equivalent to the 13 MOTUs defined in the primary run (Figure 3.2). MOTU0001 only differed once in the re-samples where sequence COI_48 split to form a singleton. MOTU0004 differed in 19 of the re-samples. In eleven of these cases, this was a result of sequence COI_57 forming a singleton MOTU, although sequences COI_84 and COI_93 also formed singletons once and twice respectively. In five of the re-samples, MOTU0004 split to form two MOTUs with differing members.

Figure 3.1 Mean number of MOTUs and standard deviation calculated from 100 re-samples at each cut-off for congruent COI, LSU and SSU gene sets (*n* = 48).

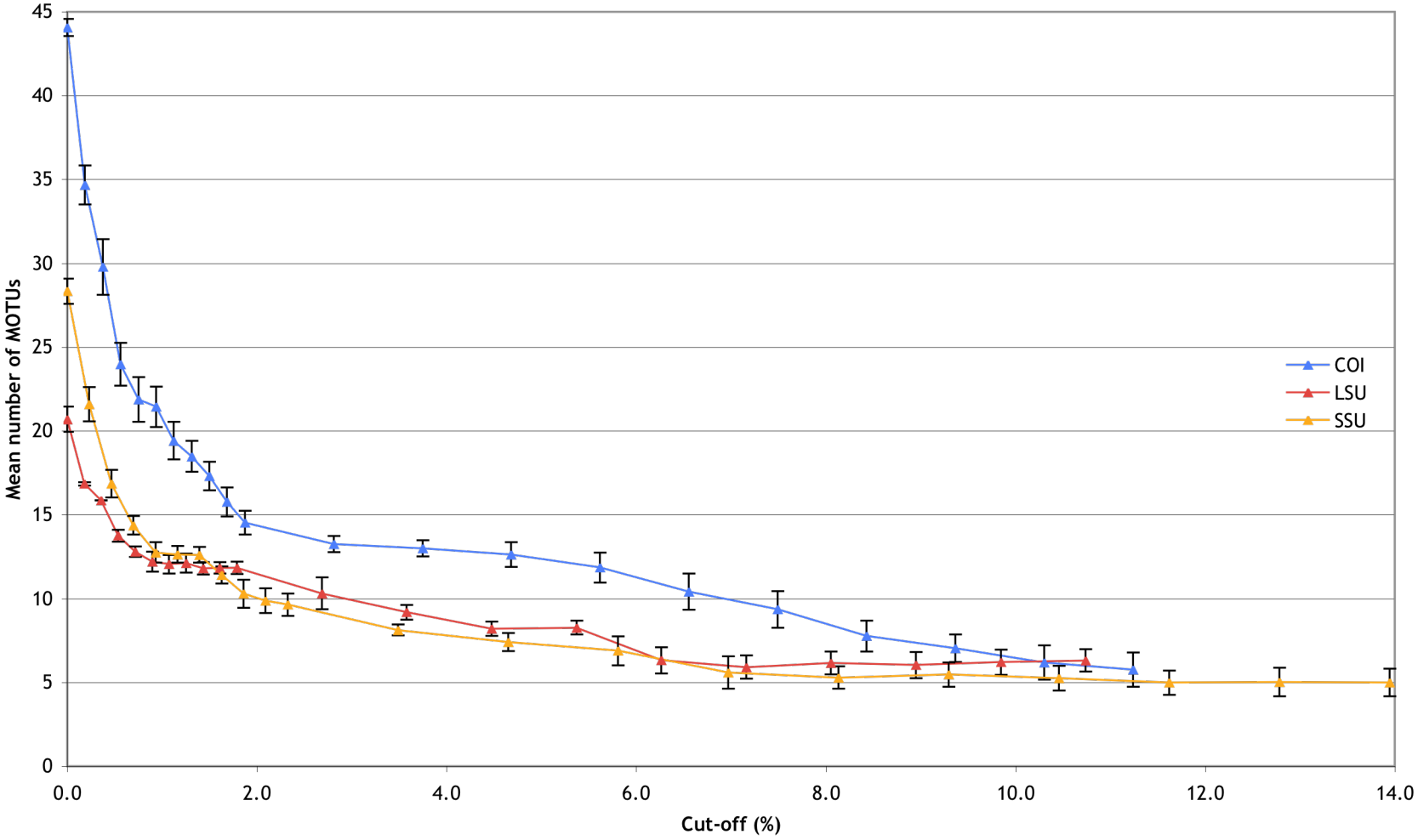
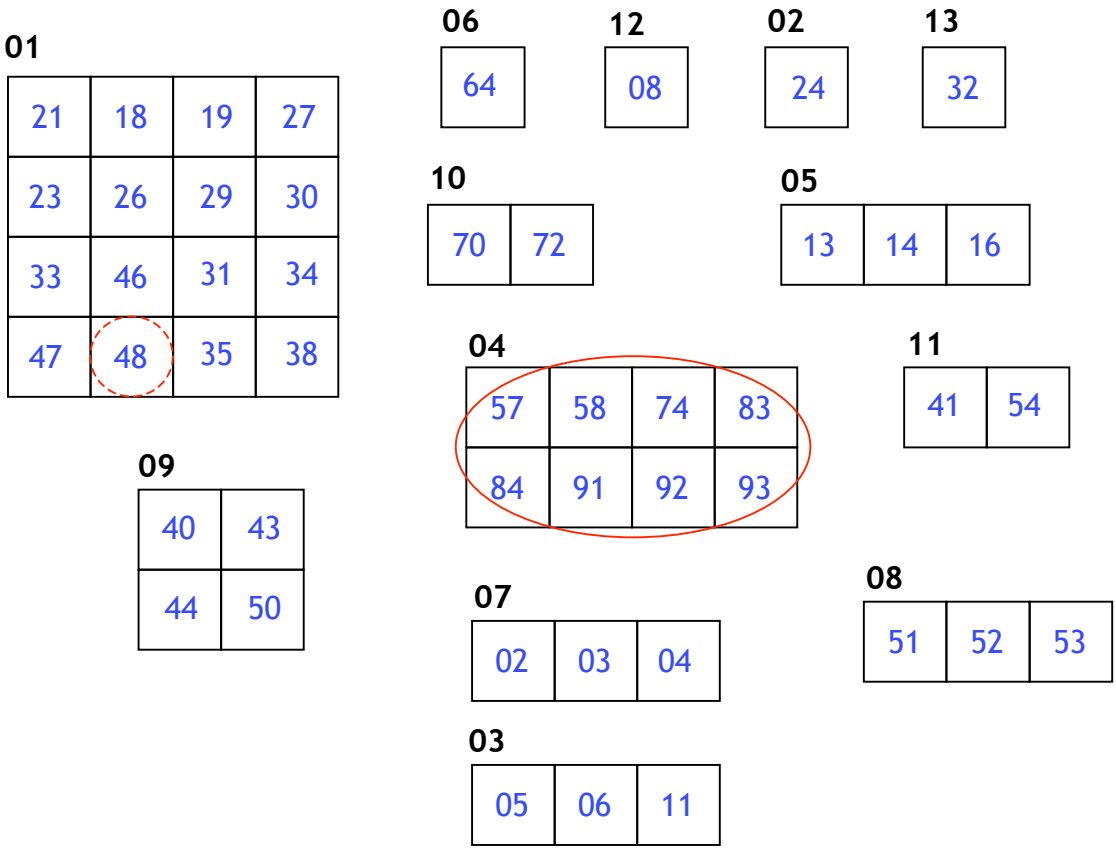


Figure 3.2 Diagram showing the members of 13 COI MOTUs using a cut-off of 15 bp from 100 re-samples. Blue numbers in boxes represent COI sequence names, black numbers represent the MOTU designation of the primary run (e.g. 01 is MOTU0001). Broken red circles indicate sequences which split from the MOTU in one of the re-samples (sequence COI_48). The solid red line highlights MOTU0004 which was only recovered in 81 of 100 re-samples. All other MOTUs were recovered in at least 99 of the re-samples.



MOTUs defined using a 15 bp cut-off were also strongly supported when a 20 bp cut-off was used. The 20 bp cut-off equivalent of MOTU0004 (Figure 3.2) differed in 14 of the 100 re-samples. The MOTU split simply in 11 of the re-samples and merged with two other MOTUs in three of the re-samples. There was one case where the MOTU had split and one of the MOTUs was joined with four other sequences. The cut-off values of 15 and 20 bp represent the range of thresholds at which the designation of sequences to and the definition of COI MOTUs was stable. This is analogous to the barcoding gap (Meyer and Paulay, 2005).

Both the LSU and SSU datasets showed a similar pattern of MOTU definition. That is both genes followed the pattern of COI, where initially there was a sharp decrease, followed by a levelling off of the number of MOTU defined, followed by another decrease.

For the LSU, the unique number of sequences (i.e. MOTUs defined at 0 bp cut-off) was 20.69±0.76 and there were 16.85±0.1 MOTUs that differed by 1 bp. This dropped to 6.31±0.66 MOTUs at 60 bp cut-off. The number of unique sequences in the COI dataset was more than double those in the LSU dataset, suggesting that COI has more variation than LSU. As the number of MOTUs at 60 bp was approximately the same (5.75±1.03 for COI), then the variation was due to small changes such as single point mutations rather than large-scale insertions and deletions.

Like the COI data, the LSU exhibited a range of stable MOTU definition (Figure 3.1) between the cut-off values of 5 and 10 bp. The mean number of MOTUs fell from 12.19±0.6 to 11.83±0.36 over 6 bp. At 8, 9 and 10 bp, where the mean number of MOTUs was 11.8 with standard deviations of 0.37, 0.35 and 0.36 respectively, the members of the MOTUs were investigated (Figure 3.3).

The majority of the re-sample MOTUs were equivalent to those designated in the primary LSU run. Differences were due to MOTUs joining. MOTU0002 merged with MOTU0008 in four of the re-samples and with MOTU0006 in 12 of the re-samples.

The LSU sequences also clustered in the same way as the COI sequences (Figures 3.2 and 3.3). For example, LSU_8bp_MOTU0003 was equivalent to COI_15bp_MOTU0010. The difference in the number of MOTUs designated between COI_15bp and LSU_8bp was due to LSU equivalents of COI_15bp_MOTU0007 (sequences 02, 03 and 04), and COI_15bp_MOTU0003 (05, 06 and 11) forming LSU_8bp_MOTU0006 containing all six sequences. A cut-off value of 30 bp was required before these six COI sequences were defined as a robust MOTU.

At first glance, the results for the SSU data appeared similar to COI and LSU. There were 28.34±0.75 unique MOTUs and the number of MOTUs decreased as the cut-off value was increased in a similar fashion to COI and LSU, with the same sharp initial decreases levelling off (Figure 3.1).

At the plateau phase, between 4 and 6 bp, the mean number of MOTUs ranged from 12.75±0.61 to 12.60±0.47. As with COI and LSU, most SSU MOTUs were stable at (Figure 3.4). Only MOTU0009 and MOTU0003 were weakly supported. The re-samples showed that the three sequences were just as likely to form one MOTU or a MOTU of SSU_51 and SSU_52 or SSU_52 and SSU_53 and a singleton of SSU_53 or SSU_51.

However, although the number of MOTUs defined at the plateau phase was the same for COI and SSU, the membership of the MOTUs was not equivalent. The COI data (and to a certain extent the LSU data) defined MOTU0001, MOTU0006 and MOTU0009 at 15 bp cut-off (Figure 3.2). In the SSU results, the member sequences of these MOTU were clustered very differently (shaded boxes, Figure 3.4). This was not due to experimental error as for some specimens, where independently derived sequences were available, the same clustering was observed (Table 3.1).

Figure 3.3 Diagram showing the members of 12 LSU MOTUs using a cut-off of 8 bp from 100 re-samples. Red numbers in boxes represent LSU sequence names, black numbers represent the MOTU designation of the primary run (e.g. 01 is MOTU0001). Arrows indicate the number of times MOTU0002 joined with MOTU0008 and MOTU0006 in the re-samples. All other MOTUs were found to be equivalent in all of the re-samples.

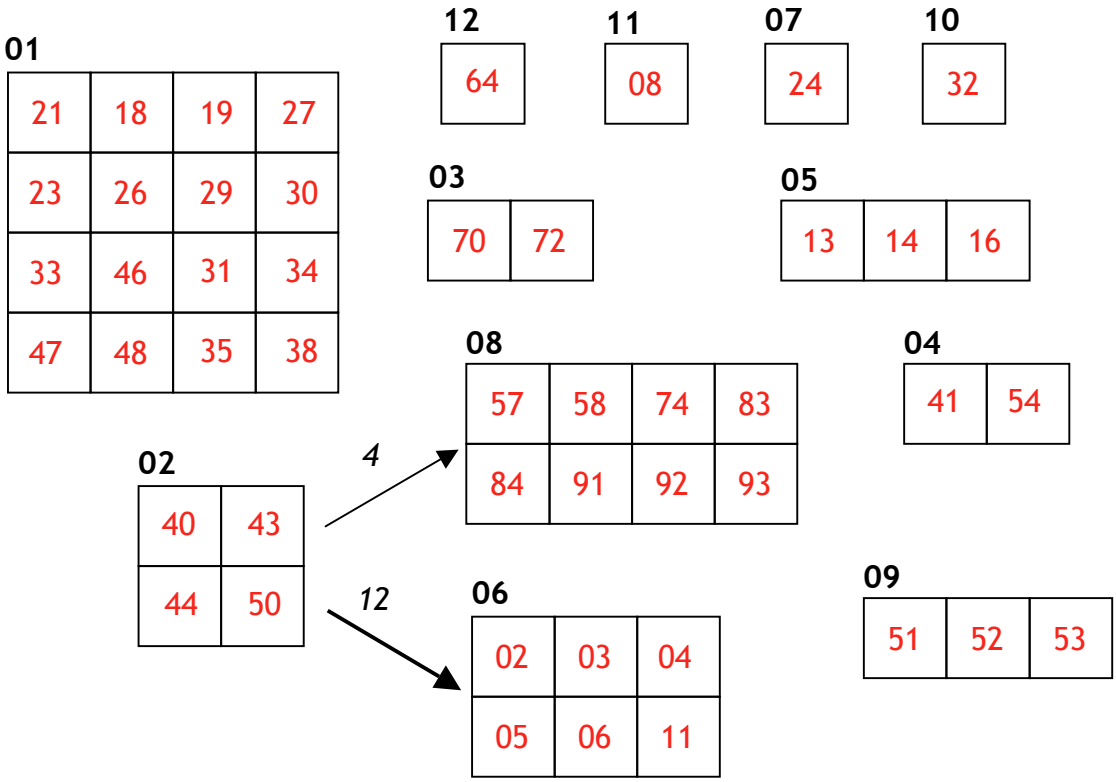
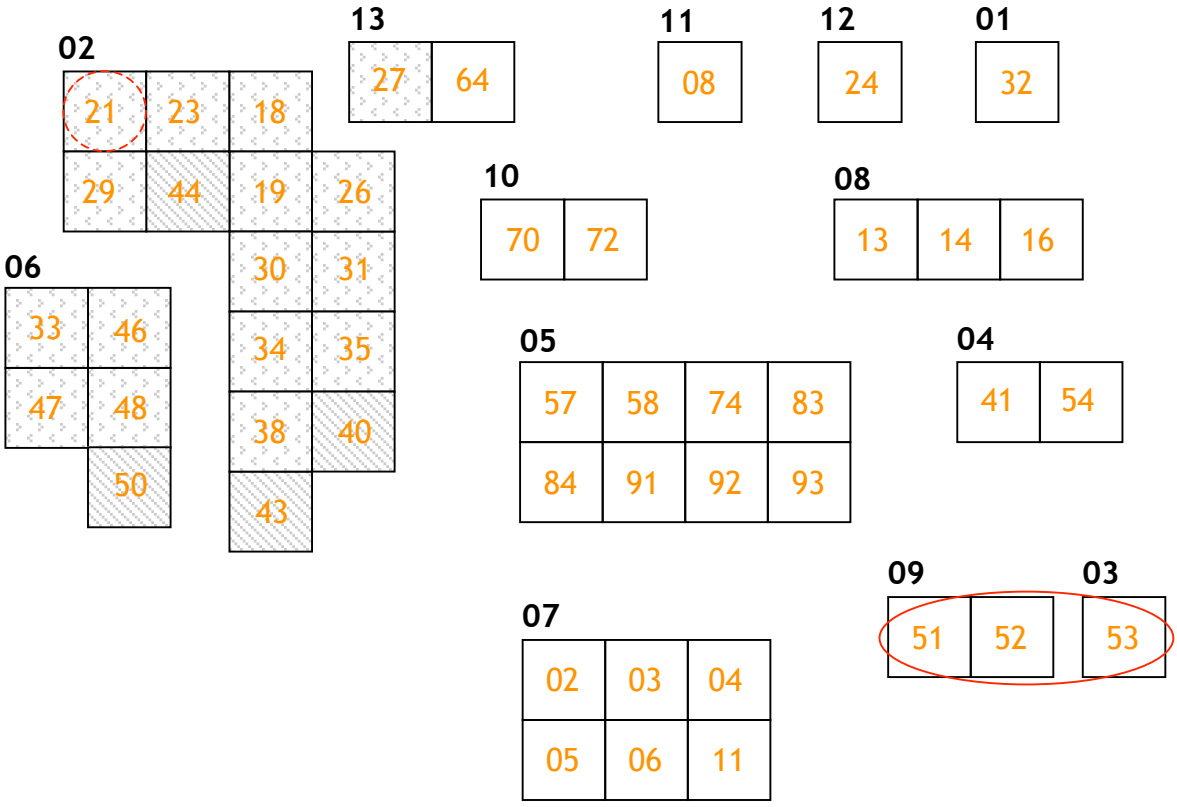


Figure 3.4 Diagram showing the members of 13 SSU MOTUs using a cut-off of 5 bp from 100 re-samples. Orange numbers in boxes represent SSU sequence names, black numbers represent the MOTU designation of the primary run (e.g. 01 is MOTU0001). Broken red circles indicate sequences that split from the MOTU in three of the re-samples (sequence SSU_21). The solid red line highlights MOTU0009 and MOTU0003 that were only recovered in 34 of 100 re-samples. All other MOTUs were recovered in at least 99 of the re-samples. Boxes with the same shading indicate the equivalent COI sequence MOTU designation.



3.3.3 DOTUR results

DOTUR (Schloss and Handelsman, 2005) was used to analyse the three gene sets for comparison with MOTU_define.pl results (Figures 3.5 - 3.7). There were only minor differences in the number of DOTUs defined using either Kimura "2-paramter" or Jukes-Cantor distances (Appendix 3.1). DOTUs generated with Jukes-Cantor distances were compared with MOTU results.

For the COI dataset, FN clustering defined more OTUs than NN clustering as expected (Figure 3.5). At 0% cut-off, there were 45 furthest neighbour DOTUs (fn-DOTUs) and 44 nearest neighbour DOTUs (nn-DOTUs). As with MOTU results, the number of fn-DOTUs and nn-DOTUs decreased rapidly as the cut-off increased. Unlike MOTU, DOTUR does not generate a plateau phase. Cut-offs were only reported where there was a change in the number of DOTUs. The decrease in the number of DOTUs slowed after 3% up to 12% cut-off where NN clustering had been used and 15% for FN clustering. The number of nn-DOTUs and fn-DOTUs deceased by three over 1% of the mean sequence length at the respective cut-offs (Figure 3.5).

There were more COI fn-DOTUs than COI MOTUs. At higher cut-offs (above 5.5%), there were more nn-DOTUs than MOTUs. Below this threshold, NN clustering defined similar numbers of DOTUs to MOTUs.

In comparison with COI DOTUR results, there were fewer LSU DOTUs found (Figure 3.6). There were 24 fn-DOTUs and 23 nn-DOTUs at 0%. The number of DOTUs fell rapidly as the cut-off was increased to 2%. Unlike the COI DOTUs (and SSU), the number of fn-DOTUs and nn-DOTUs were very similar (Figure 3.6). Again, there was no plateau phase. However, between 3.2-8.5% and 4.1-8.9% (for nn-DOTUs and fn-DOTUs respectively) the number of DOTUs for both clustering methods decreased by one.

Figure 3.5 The number of COI OTUs defined using furthest (FN) and nearest (NN) neighbour clustering in DOTUR and MOTU_define.pl for the congruent COI dataset at each cut-off as a percentage of the mean sequence length.

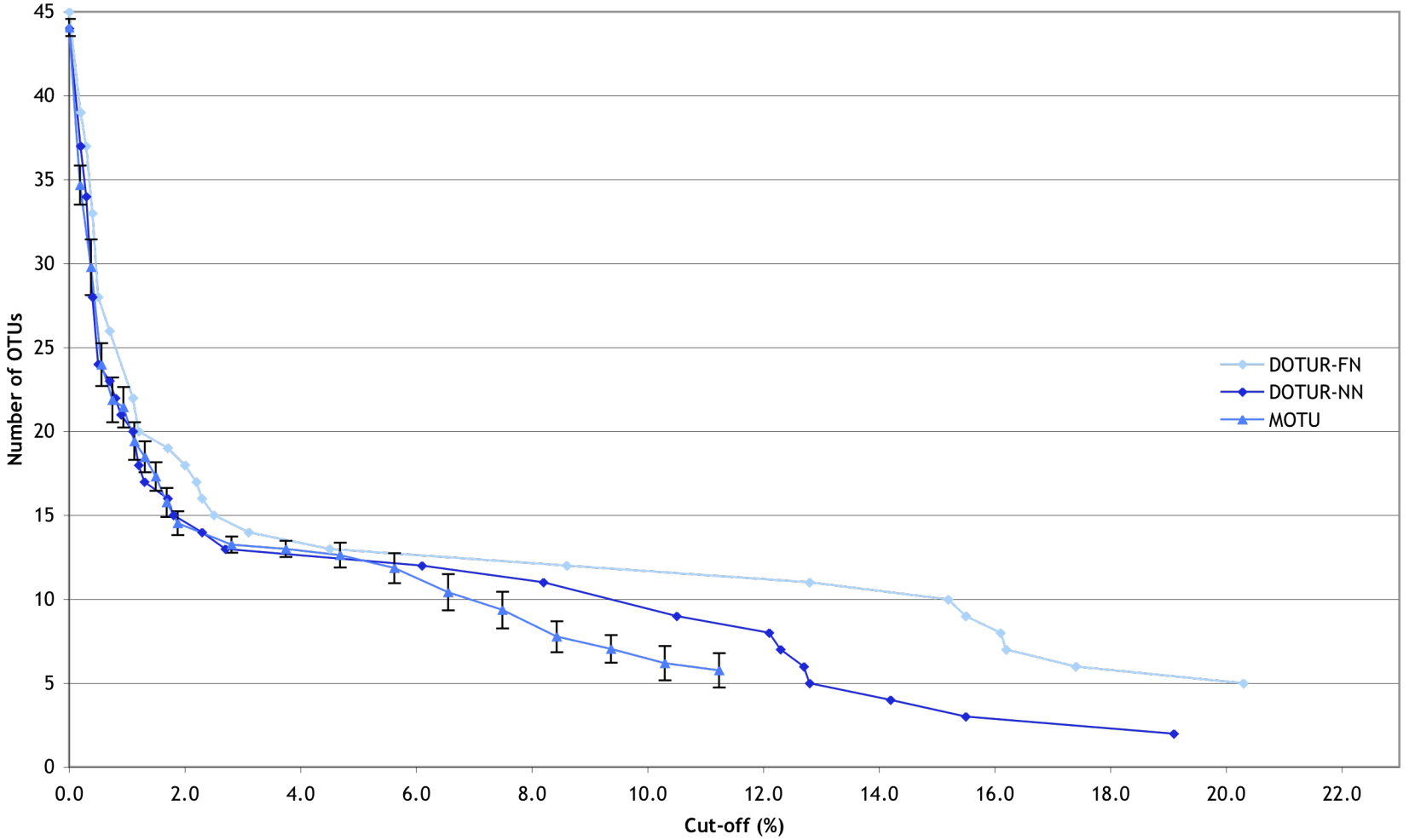


Figure 3.6 Number of LSU OTUs defined using FN and NN clustering in DOTUR and MOTU_define.pl for the congruent LSU dataset at each cut-off as a percentage of the mean sequence length.

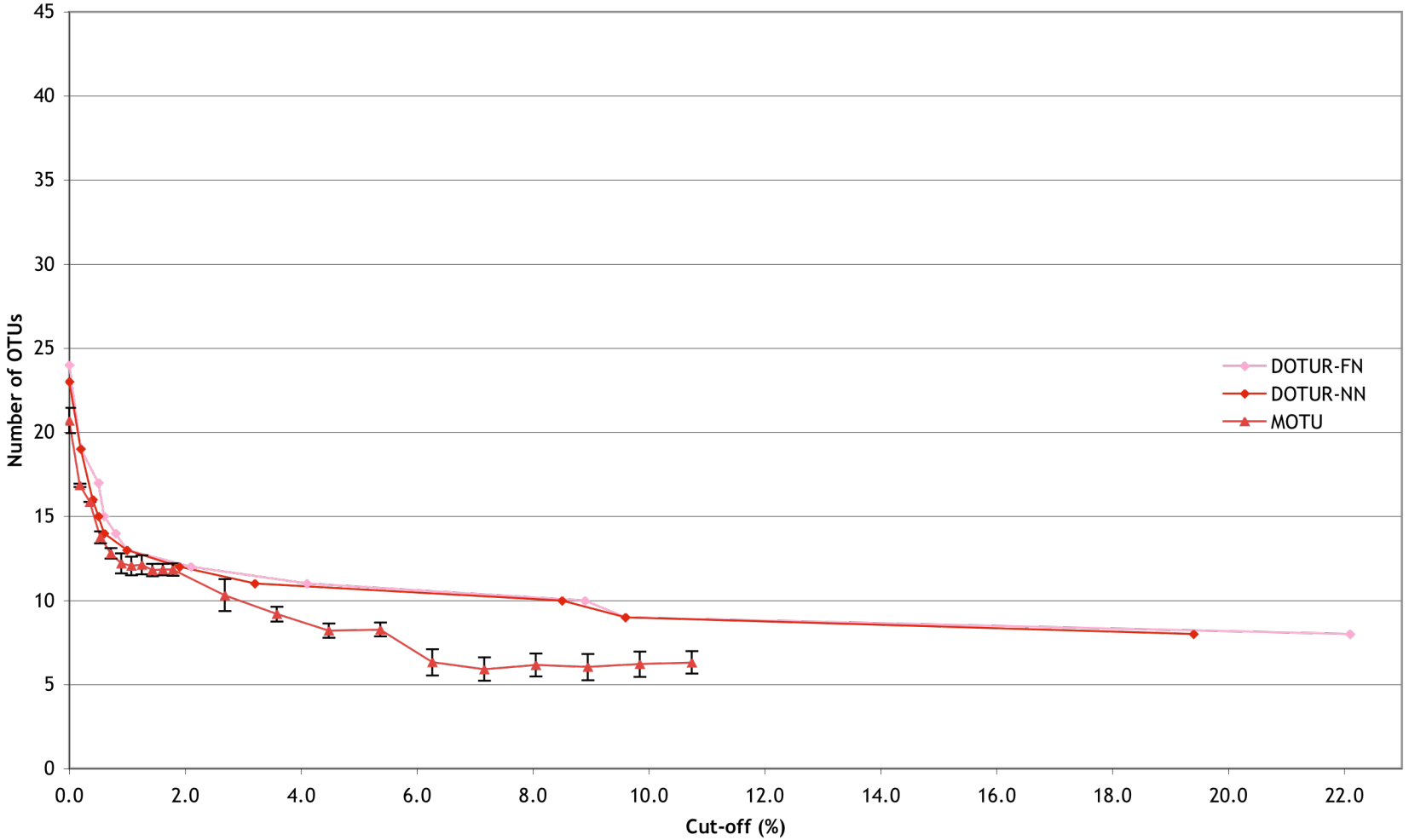
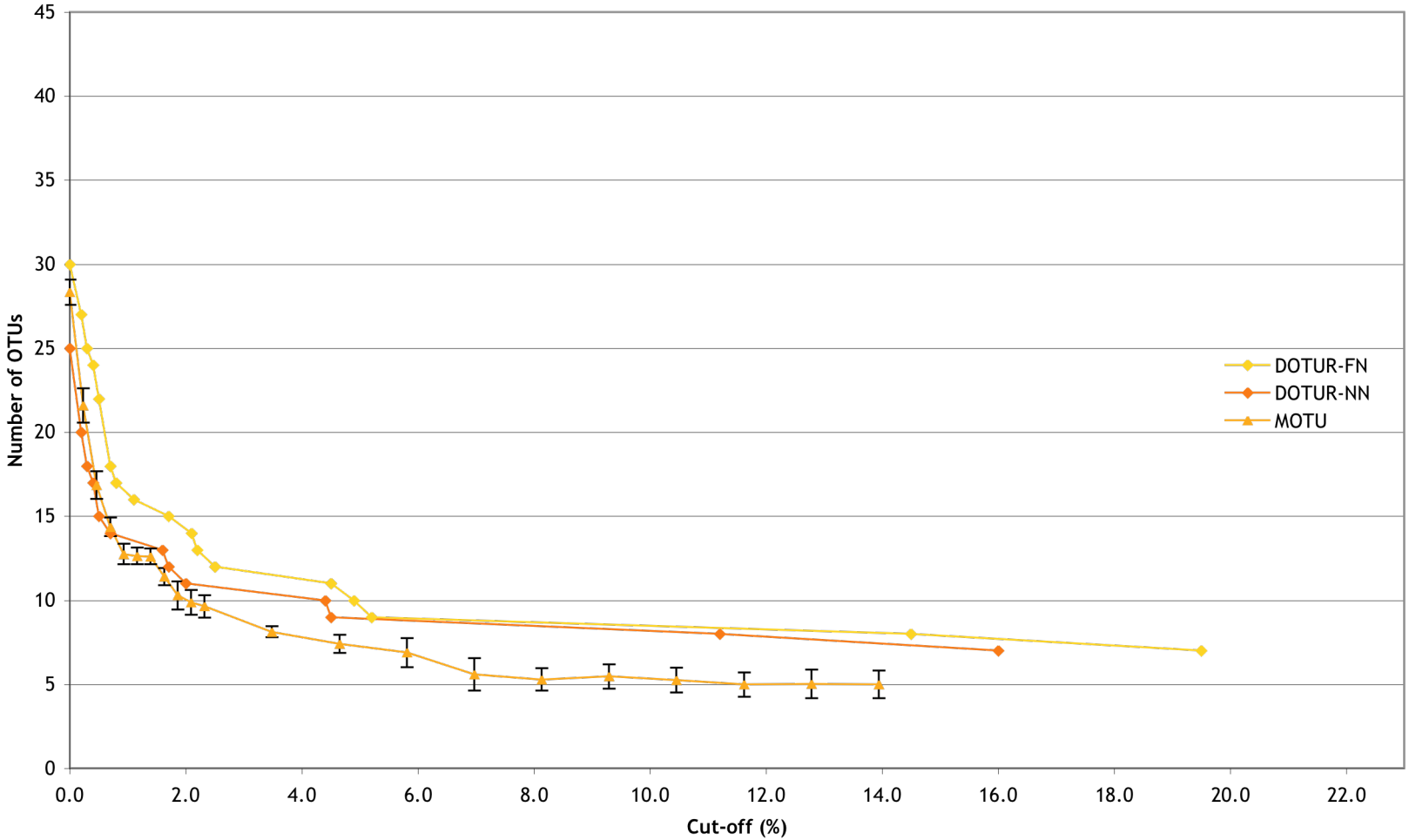


Figure 3.7 Number of SSU OTUs defined using FN and NN clustering in DOTUR and MOTU_define.pl for the congruent SSU dataset at each cut-off as a percentage of the mean sequence length.



At 0% cut-off, there were 30 fn-DOTUs and 25 nn-DOTUs (Figure 3.7). This was the only dataset where there were more SSU MOTUs found at 0% cut-off than nn-DOTUs. There was a rapid reduction in the number of DOTUs until 1% cut-off. This initial decrease phase was much shorter in the SSU results compared with the COI and LSU results. Unlike COI and LSU, the SSU data exhibited two phases where the proportional decrease (i.e. where a small change in the number of DOTUs occurred over a large range of cut-off values) in the number of DOTUs over the range of cut-offs was small for both fn-DOTUs and nn-DOTUs (Figure 3.7).

Membership of sequences to DOTUs was investigated where the proportional decrease in the number of DOTUs was smallest for FN and NN clustering. For COI and LSU DOTUs, NN and FN clustering had only one phase where the number of DOTUs decreased by one over a range of cut-off values (Figure 3.8 and 3.9). The SSU data indicated two phases, so membership of sequences to DOTUs at both phases was investigated (Figure 3.10).

As with MOTUs, COI and LSU DOTUs are congruent at the beginning of the phase but cluster differently. Not only are SSU DOTUs different to COI and LSU DOTUs, but NN and FN clustering also result in different orders and defined different DOTUs (Figure 3.10).

Figure 3.8 Diagram to show how COI DOTUs change over the smallest proportional decrease. For NN clustering this was at 2.7 - 6.1%. For FN clustering there were two phases, 4.5 - 8.6% and 8.6 - 12.8%. Broken line shows which of the nn-DOTUs and the first phase of fn-DOTUs merged at the different cut-offs. The double-headed arrow indicates which fn-DOTUs merged at the second phase.

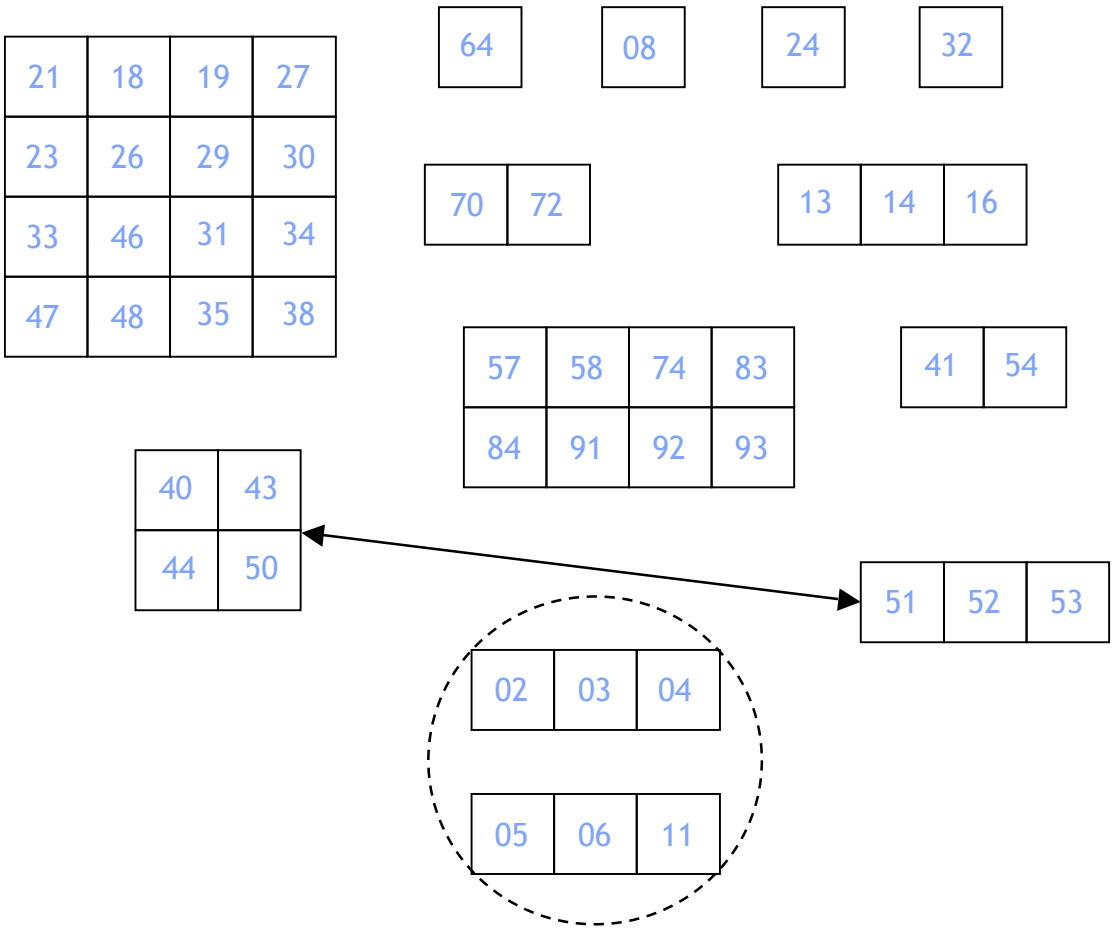


Figure 3.9 For LSU DOTUs the smallest proportional change was between 3.2 - 8.5% for NN clustering and 4.1 - 8.9% for FN clustering. The double-headed arrow indicates the merging DOTUs, which were the same for NN and FN clustering. The star highlights the sequences that formed two MOTUs at the plateau phase.

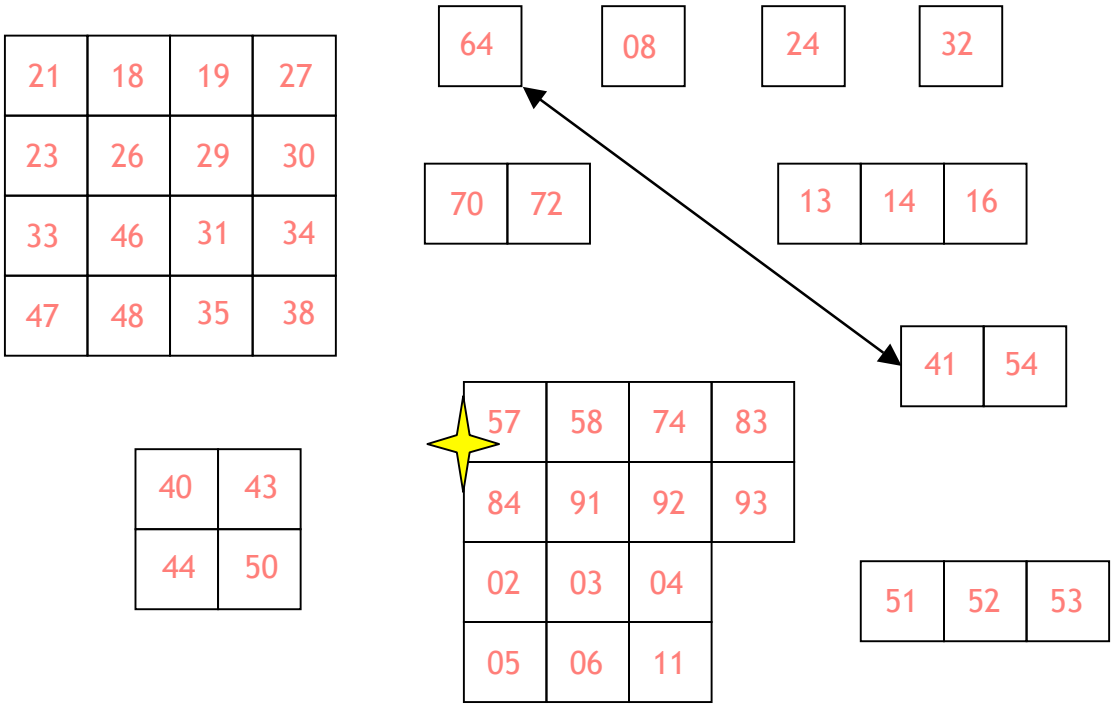
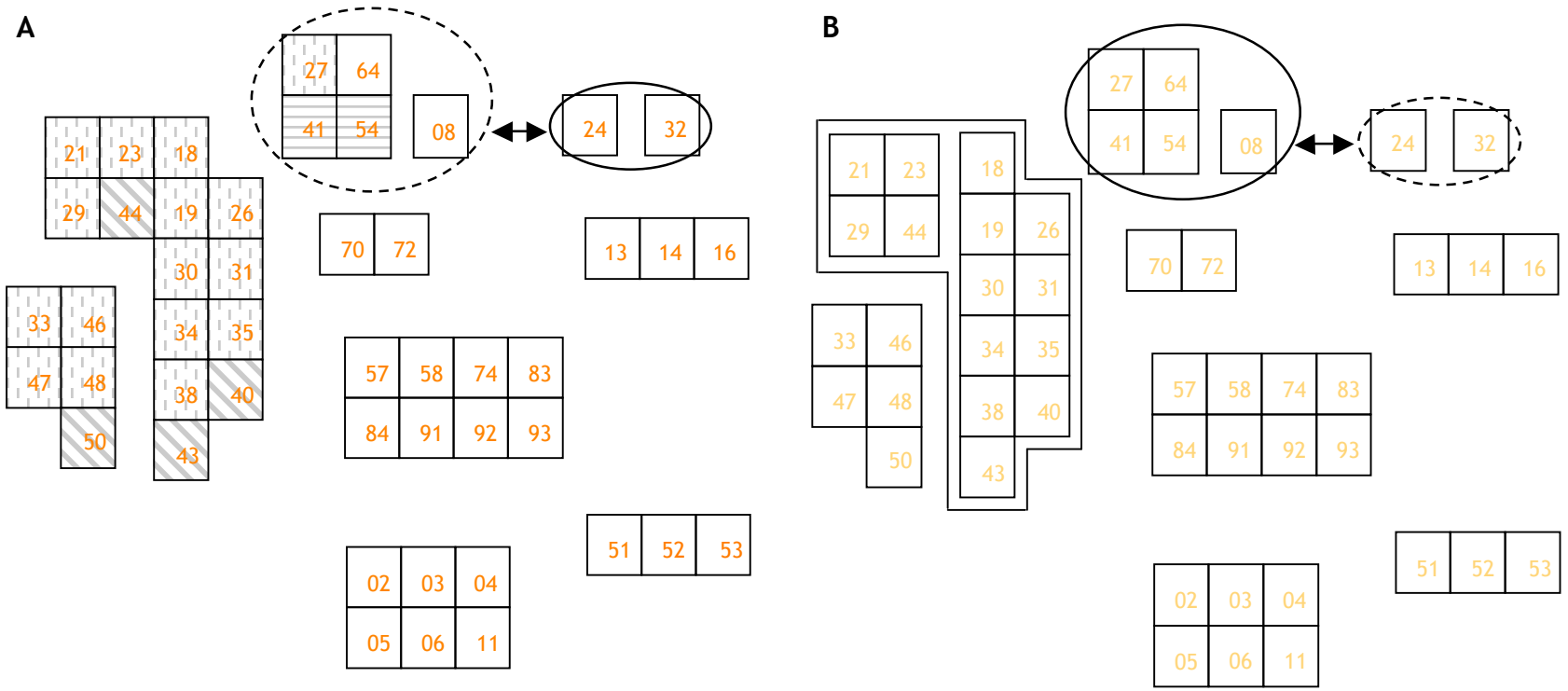


Figure 3.10 Diagram to show how SSU DOTUs change at both phases of smallest proportional decrease. **A** represents changes in NN clustering, the broken line shows where nn-DOTUs changed between 2% and 4.4% (first phase) from 11 to 10 DOTUs. Sequences SSU_24 and SSU_32 join before the second phase (solid line). The double headed arrow shows where nn-DOTUs changed in the second phase (4.5% to 11.2%). Shaded boxes represent the equivalent COI sequence DOTU designations. **B** represents changes in FN clustering. In the first phase (2.5% to 4.5%) sequences SSU_24 and SSU_32 join (dashed line). Then two fn-DOTUs join (solid line) before the second phase (double headed arrow).



3.3.4 MOTUs versus DOTUs

The clustering behaviour of COI sequences were broadly similar between MOTU_define.pl and DOTUR. The main difference was DOTUR grouped sequences from MOTU0008 and MOTU0009 at the phase of smallest proportional change (Figures 3.2 and 3.8). This did not occur in the MOTU results until a 35 bp cut-off was used, and then the sequences were members of a MOTU containing many more sequences.

The first difference between LSU MOTUs and DOTUs, was the definition of one DOTU in comparison with two MOTUs (starred DOTU, Figure 3.9). This was where the number of OTUs was most stable. The merged MOTU was not supported by MOTU_define.pl until a cut-off of 20 bp (3.6%) was used. The second difference in LSU DOTUs was the joining of sequences LSU_41 and LSU_54 with LSU_64. MOTU_define.pl did not support this join. Sequence LSU_64 joined with LSU_08, LSU_24 and LSU_32 before LSU_41 and LSU_54 at 35 bp cut-off.

Overall, the SSU MOTUs defined were similar to either SSU nn-DOTUs or fn-DOTUs. Again, SSU OTUs did not form equivalent OTUs in comparison with the COI or LSU results.

3.3.5 Tree comparisons of all sequences

Unrooted NJ phylograms were constructed for the complete gene sets (Figures 3.11, 3.12 and 3.13). Branch tips were compared with mean MOTU designation from the plateau phase of each gene.

Figure 3.11 Unrooted NJ phylogram of complete COI data set. Blue sequence names are from the congruent data set and have the MOTU ID number at a 15 bp cut-off (from Figure 3.2).

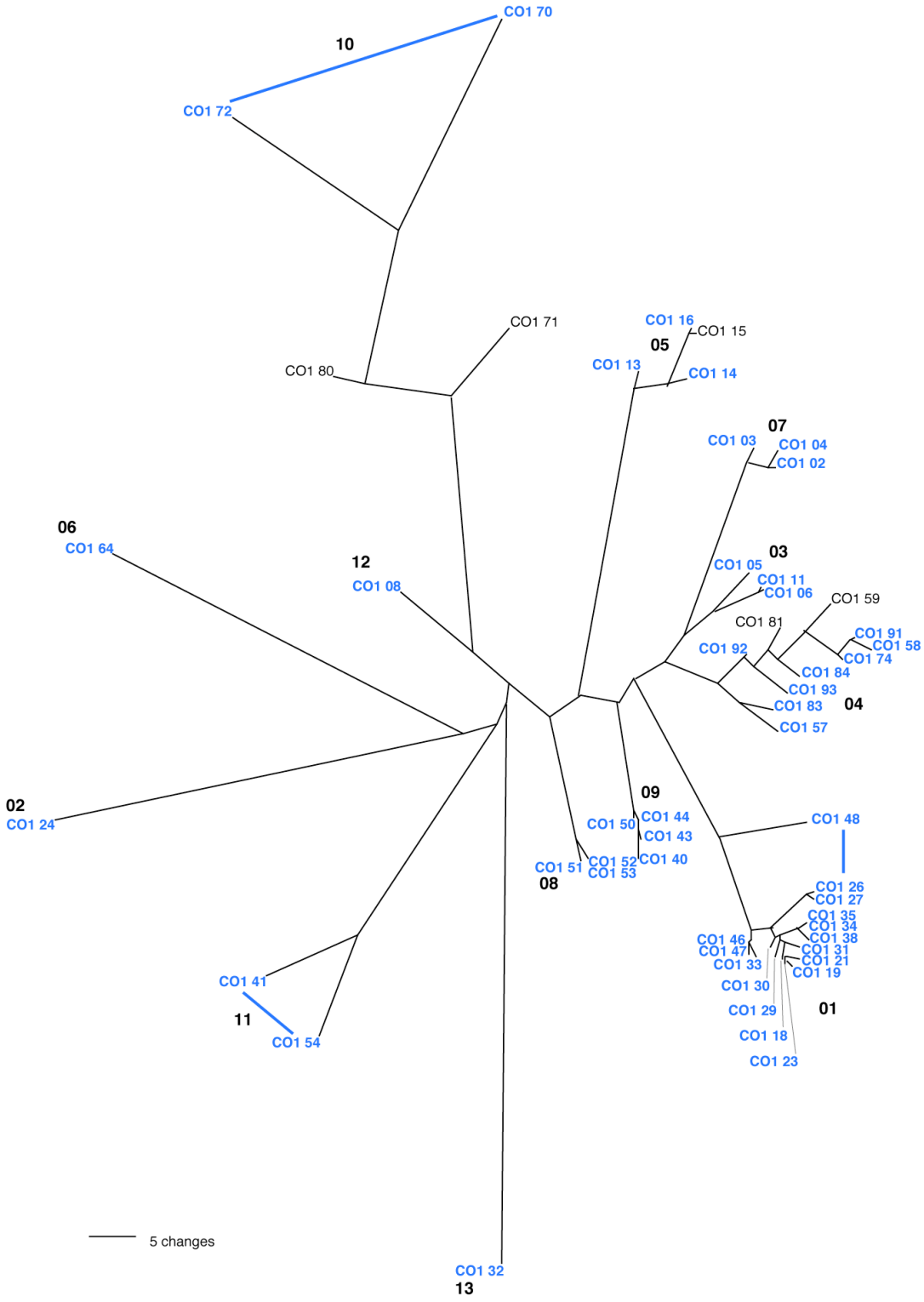


Figure 3.12 Unrooted NJ phylogram for complete LSU data set. Red sequence names are from the congruent data set and have the MOTU ID number from the 8 bp cut-off run (Figure 3.3). Solid lines and number represent corresponding MOTUs defined using 8 bp cut-off from Figure 3.3. Italicised sequence names segregate into two clusters in the LSU results (cluster 01, blue arrow and cluster 02, purple arrow), but are mixed in the SSU clusters (Figure 3.13).

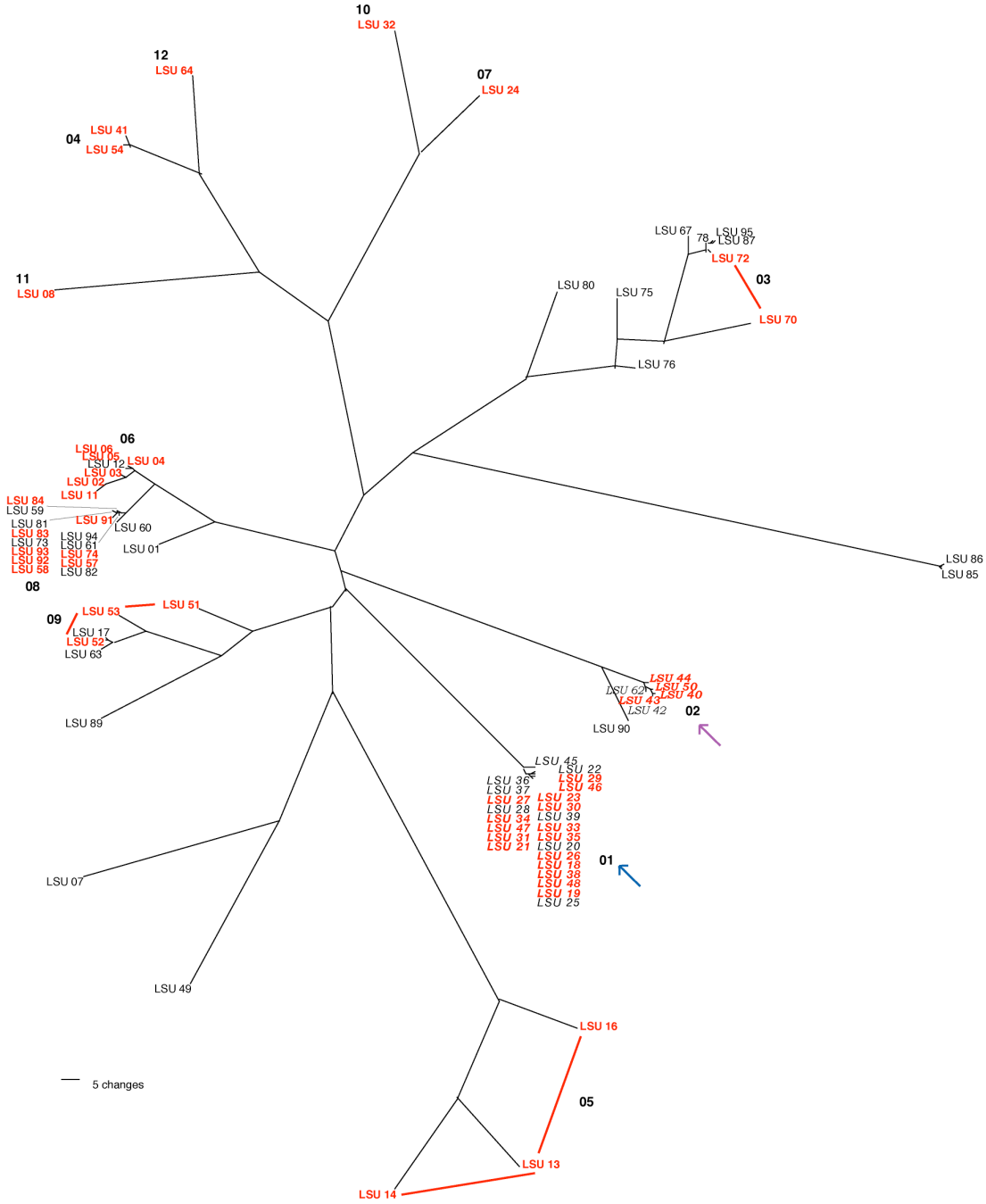
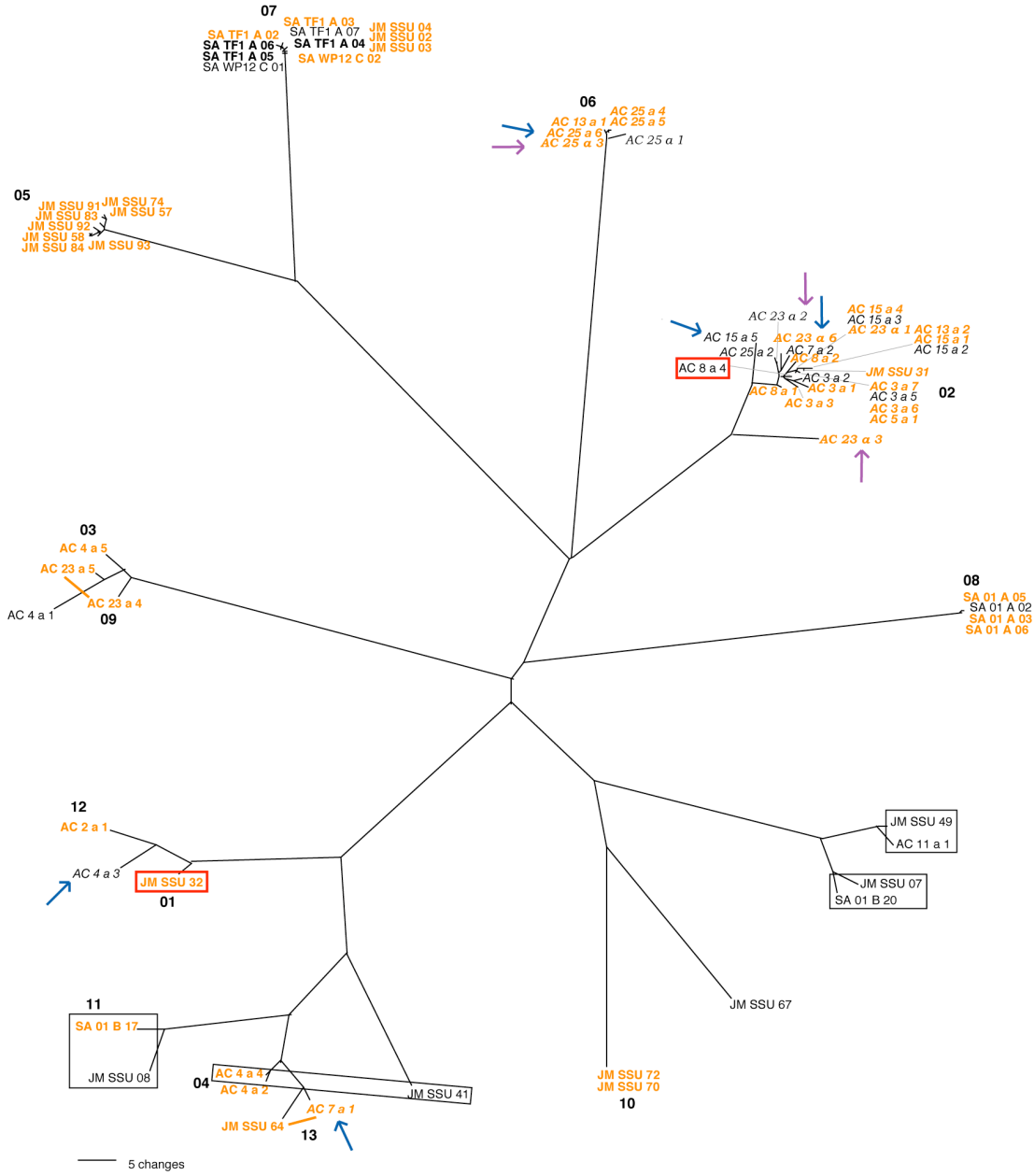


Figure 3.13 Unrooted phylogram of complete SSU data set. Orange sequence names are from the congruent data set and have MOTU ID numbers from Figure 3.4. Sequence pairs in black boxes indicate duplicates sequences, red boxes highlight duplicated sequences which do not cluster together in the phylogram. Bold black sequence names in the cluster comparable to MOTU 07, are also duplicated sequences. Blue arrows highlight SSU clusters which contain sequences from LSU cluster 01 and purple arrows show SSU clusters which contain sequences from LSU cluster 02 in Figure 3.12.



The addition of the extra sequences to each gene set did not alter the COI and LSU MOTUs and there was similarity in branching patterns (Figures 3.11 and 3.12). There were 48 sequences in the congruent data set for each gene, so the addition of five COI sequences may not be expected to dramatically change the COI MOTU designations. However, the LSU data set was increased from 48 to 82 sequences, yet the clusters did not change.

As seen in the MOTU and DOTUR results, SSU clusters differed from COI and LSU clusters (Figure 3.13). COI and LSU clusters separate sequences 70 and 72, where as for the SSU results show that the sequences are the same. Most SSU pairs of duplicated sequences were found to cluster close together (grey sequence names, Figure 3.13). There are two exceptions. The first is JM_SSU_41 and AC_4-a-4 which were generated from the same specimen, but are separated on the tree (boxed sequence names, Figure 3.13). JM_SSU_41 is 316 bases compared to 457 bp for AC_4-a-4 and 459 bases of AC_4-a-2 (Appendix 3.2). Since the tree is based on absolute differences including end gaps, significant differences in length are likely to separate sequences on the phylogram. The second, even more striking exception, were sequences JM_SSU_32 and AC_8-a-4 (red boxed sequence names, Figure 3.13). These sequences were generated using the same primers, sequenced using the same primer, are 416 and 382 bases respectively, yet are on opposite branches. The plausible explanation for this surprising pattern is due to mislabelling of this particular specimen. Great care was taken to ensure correct specimen order during amplification and sequencing. However there was no monitoring of specimens before they were received as lysis samples in a 96-well plate. Without being able to trace the complete history of individual samples, this needs to be considered as the most likely cause of this result.

3.4 Discussion

3.4.1 PCR and Sequencing success

The results clearly indicate that LSU is more reliably recovered from this sample set than SSU and especially COI sequences. Differences among the number of positives were not due to a lack of DNA as PCRs were successful for the other genes. There was no systematic bias of the recovery COI sequences. Comparing LSU taxonomic designations (from BLAST hits), there was no relationship between the sub-order of the LSU sequence and a positive COI PCR (Appendix 3.2). If COI barcodes are not universally recoverable from all specimens, then there is little to be gained by basing barcode system on COI. Specific primers are sometimes required to amplify COI depending on the taxa, for example spider mites (Ros and Breeuwer, 2007). If this is found to be true across many taxa, then using COI as a standard barcode would be unsuitable, where the specimen to be analysed had not already been classified to a taxonomic group. The LSU primers worked extremely well with a sequence generation success rate of over 90%. These primers have also worked well on other taxa (e.g. tardigrades, copepods (Mann, unpublished)). If a specimen was rare then it would be more appropriate to use a target that has a high probability of being recovered with PCR. Although this is not so much a problem for large organisms, which can be re-sampled easily, when the whole animal is degraded for lysis, as in meiofaunal barcoding, there is a limit to the number of PCRs one is able to perform. This may also be the case for larger organisms where obtaining DNA is problematic, for example when the specimen is old and DNA is degraded, or when it is not easily re-sampled because the specimen is rare.

3.4.2 MOTU_define.pl

For the congruent dataset, there were more COI MOTUs defined than LSU and SSU MOTUs. There were only similar numbers of MOTUs for each gene,

when the cut-off value was between 10-11% (Figure 3.1). At low cut-off values (less than 1.5%), there were more SSU MOTUs than LSU. Increasing the cut-off value above 1.5% results in more LSU MOTUs than SSU MOTUs. If SSU has more MOTUs at low cut-off values, then it may be a more accurate predictor of individual variation. At higher values, the LSU barcodes exhibit more deep-seated differences, which may relate to differences between species.

3.4.3 DOTUR

Within each gene set, FN clustering generated more DOTUs than NN clustering. As with the MOTU data, there were more COI DOTUs than LSU or SSU DOTUs, up to a cut-off value of 12% and 15% for NN and FN clustering respectively. The number of COI nn-DOTUs and fn-DOTUs decreased by four OTUs over approximately 1% difference. This was a striking feature of the COI DOTUs only, and was not reproduced in either of the nuclear gene sets. In the absence of any morphological species identification, it is difficult to identify the root of this feature. It is unlikely to be sequences segregating into species groups at the high cut-off value. It is more probable to be indicative of the grouping of species into genera.

DOTUs generated from the LSU data exhibited minor differences between the FN and NN clustering and the membership of sequences to DOTUs was very similar. The decrease in the number of LSU DOTUs was smaller, and occurred over a larger range than COI DOTUs. This would indicate that different LSU lineages were more isolated than the COI sequences. If LSU lineages were similar, the number of DOTUs would be expected to fall more sharply as they would be assigned to the same DOTU at lower cut-off values.

Whereas LSU DOTUs were similar both in the number generated and in the membership of sequences to a particular DOTU, SSU DOTUs differed in both aspects. FN clustering defined an average of 5 more DOTUs than NN clustering up until a 2% cut-off. This difference was greater than that seen between COI and LSU nn-DOTUs and fn-DOTUs. Moreover, membership of sequences differed in nn-DOTUs when compared to fn-DOTUs. It was

expected to have more fn-DOTUs than nn-DOTUs due to the stringent rules for assigning sequences. It was also expected that LSU and SSU would have similarities as both are nuclear genes, therefore subject to the same selection pressure, and are inherited in tandem.

3.4.4 MOTU vs. DOTUR

Results of MOTU and DOTUR were generally similar. The number of DOTUs for each gene set was more than the corresponding MOTUs. These differences are probably due to the multiple versus the pairwise alignment of DOTUR and MOTU_define.pl respectively. In a multiple alignment, differences in length can lead to sequences being defined in different DOTUs. Therefore, DOTUR is likely to overestimate the number of taxa at any one cut-off.

Another issue with DOTUR is the way in which the cut-offs are reported only when the number of DOTUs changes. There is an obligate decrease in the number of DOTUs as the cut-off is increased. Where as MOTU_define.pl generates MOTUs at each cut-off, it is possible to see a plateau in the number MOTU, DOTUR does not. Working out the flattest part of the line for DOTUs could highlight a potential plateau, as during that phase the number and membership of DOTUs does not change.

Different MOTUs could be generated depending on the addition order of sequences, though running multiple re-samplings could highlight robust MOTU. As DOTUs were ultimately based on the same multiple alignment, whatever clustering method was used, DOTUR should generate the equivalent DOTUs. In COI and LSU datasets, FN and NN clustering formed equivalent DOTUs, albeit at different cut-off values. However, fn- and nn-DOTUs were incongruent in the SSU data. If one clustering method produces different OTUs from another, both using the same alignment, without any means of investigating the robustness of either result, which one is to be believed?

3.4.5 Complete data comparisons

The inclusion of all the sequences in the analysis did not alter MOTUs defined using the congruent data set. New MOTUs were generated from the additional sequences in the LSU data set (Figure 3.12).

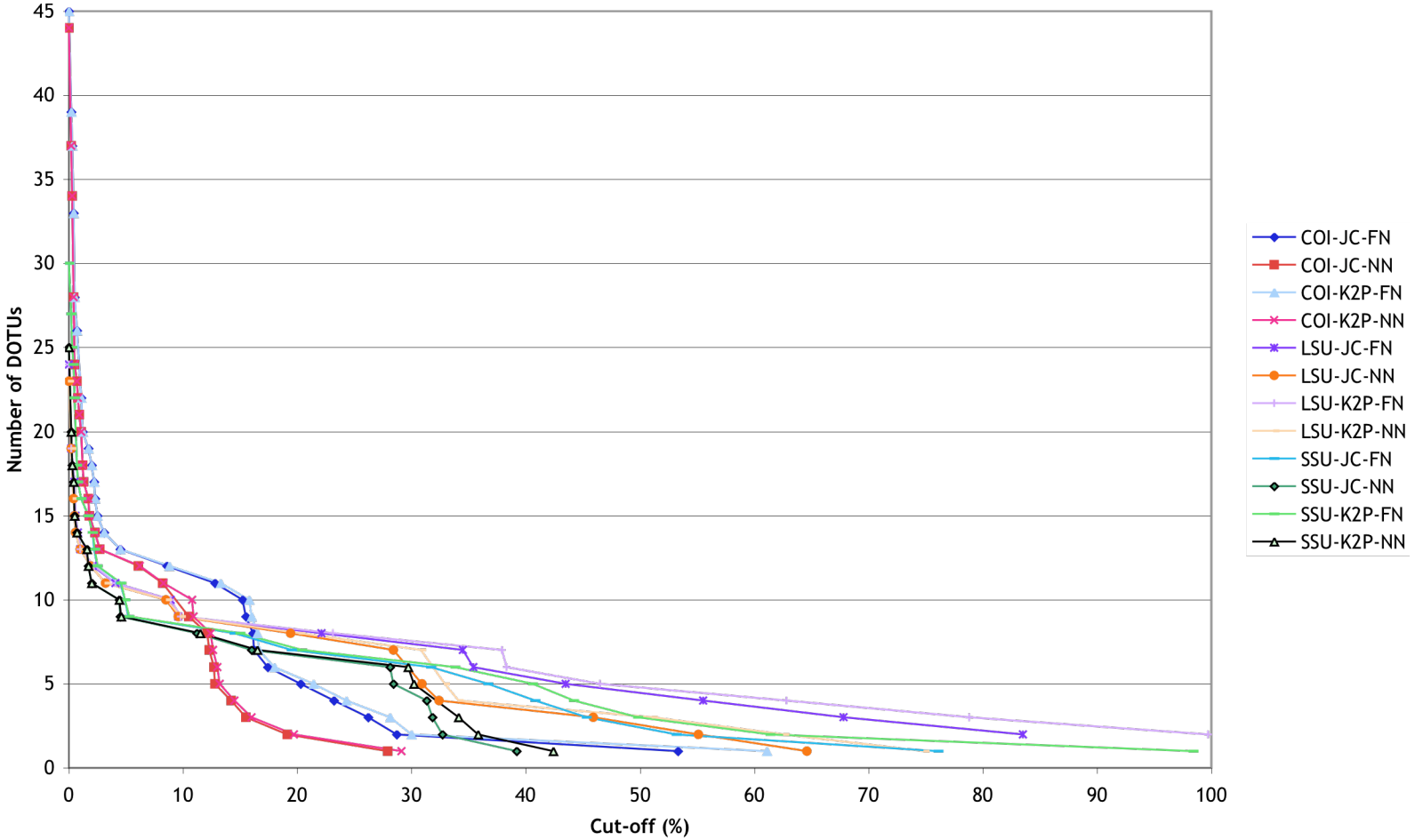
3.4.6 What gene(s) to use?

The most OTUs defined were from the COI gene set, regardless of which method is used. However, the overall sequence success rate was the lowest in this investigation. The LSU gene set had the greatest success rate. Moreover, LSU OTUs were very similar to the COI OTUs. This was unexpected as COI is a mitochondrial gene and LSU is a nuclear gene. If LSU OTUs are an accurate predictor of COI clustering behaviour, and sequences are more reliably recovered from samples, then LSU would be an improved proposition as a universal barcode target.

As demonstrated by these results, it should not be used in isolation. Whilst the COI and LSU OTUs were broadly similar, both were different from the SSU OTUs generated. It is surprising that the SSU and LSU results are incongruent (as they are found on tandem repeats), but this case highlights the necessity to consider more than one target for barcoding.

APPENDIX 3.1

Complete DOTUR results using Kimura "2-Parameter" and Jukes-Cantor distances and FN and NN clustering methods for COI, LSU and SSU, showing the number of DOTUs defined against the cut-off values as a percentage of the mean sequence length.



APPENDIX 3.2

Table detailing sequence lengths (bp), top BLAST hit for each gene (GenBank accession number), e value (E). Cells shaded grey are sequences which were not included in analysis (either because there were no sequences generated, or they were too short or they were non-nematode sequences). Italicised SSU entries represent sequences that were generated prior to this study.

Name	COI			LSU			SSU		
	bp	Top hit	E	bp	Top hit	E	bp	Top hit	E
01				236	EU303298	$3e^{-106}$	*402*	*EU196007*	*0*
02	582	U57030	0	564	EU195964	0	443	EU196007	0
	--	--	--	--	--	--	*403*	*EU196007*	*0*
03	535	U57030	0	557	EU195964	0	443	EU196007	0
	--	--	--	--	--	--	*399*	*EU196007*	*0*
04	634	U57030	0	571	EU195964	0	443	EU196007	0
	--	--	--	--	--	--	*400*	*EU196007*	*0*
05	407	AY508036	$2e^{-158}$	573	EU195964	0	402	EU196007	0
06	452	DQ285543	$4e^{-167}$	571	EU195964	0	402	EU196007	0
07	513	EF126762	$4e^{-46}$	609	DQ903104	0	523	EU543175	0
	--	--	--	--	--	--	*440*	*EU543175*	*0*
08	439	AY508036	$7e^{-145}$	644	AY294183	0	529	AY284681	0
	--	--	--	--	--	--	*455*	*AY284681*	*0*
11	446	DQ285543	$3e^{-162}$	566	EU195964	0	*400*	*EU196007*	*0*
12				565	EU195964	0	*400*	*EU196007*	*0*
13	486	U57030	$4e^{-160}$	512	AF549407	0	*433*	*AF430482*	*0*
14	498	U57030	$1e^{-159}$	636	EU195982	0	*435*	*AF430482*	*0*
15	621	U57030	0				*434*	*AF430482*	*0*
16	663	EU652745	0	425	AF549407	0	*435*	*AF430482*	*0*
17				571	EF990722	0	*460*	*U73452*	*0*
18	654	FJ483518	0	559	EU195969	0	*438*	*EU196009*	*0*
19	667	FJ483518	0	571	EU195969	0	*431*	*EU196009*	*0*
20				583	EU195969	0	*432*	*EU196009*	*0*
21	601	DQ408632	0	551	EU195969	0	*414*	*EU196009*	*0*
22				444	EU195969	0	*416*	*EU196009*	*0*
23	553	DQ408632	0	587	EU195969	0	*440*	*EU196009*	*0*
24	645	FJ483518	0	610	EF417144	$8e^{-140}$	*465*	*AM088400*	*0*
25				569	EU195969	0	*461*	*AM088400*	*0*
26	633	EU407803	0	583	EU195969	0	*415*	*EU196009*	*0*
27	600	AF303158	0	540	EU195969	0	*452*	*EF514920*	*0*
28				549	EU195969	0	*417*	*EU196009*	*0*
29	602	EU407803	0	557	EU195969	0	*379*	*EU196009*	*0*
30	469	EU407803	$7e^{-164}$	584	EU195969	0	*339*	*EU196009*	*$2e^{-170}$*
31	625	EU407803	0	553	EU195969	0	446	EU196009	0
32	619	EU407788	$8e^{-163}$	678	EF417144	$2e^{-148}$	416	AM088400	$2e^{-175}$
	--	--	--	--	--	--	*382*	*EU196009*	*0*
33	634	EU652745	0	584	EU195969	0	*433*	*U13929*	*0*
34	627	FJ483518	0	550	EU195969	0	*431*	*EU196009*	*0*
35	643	FJ483518	0	569	EU195969	0	*431*	*EU196009*	*0*
36				571	EU195969	0	*430*	*EU196009*	*0*
37				566	EU195969	0	*431*	*EU196009*	*0*
38	629	FJ483518	0	566	EU195969	0	*431*	*EU196009*	*0*
39				583	EU195969	0	*326*	*EU196009*	*0*
40	459	EU407799	0	623	AY604481	0	*431*	*EU196009*	*0*
41	440	EU407803	$1e^{-141}$	584	EF417146	0	316	EF514920	$3e^{-154}$
	--	--	--	--	--	--	*457*	*EF514920*	*0*

42	627			576	AY604481	0	*426*	*EU196006*	*0*
43	459	EU407799	0	572	AY604481	0	*338*	*EU196006*	*$5e^{-170}$*
44	457	EU407803	0	495	EF417140	0	*379*	*EU196006*	*0*
45				490	EU195969	0	*417*	*EU196006*	*0*
46	529	EU407803	0	557	EU195969	0	*435*	*U13929*	*0*
47	569	DQ408632	0	552	EU195969	0	*435*	*U13929*	*0*
48	332	EU407803	$4e^{-110}$	570	EU195969	0	*434*	*U13929*	*0*
49				476	DQ903079	$4e^{-137}$	474	AF202161	0
	--	--	--	--	--	--	*514*	*EU543175*	*0*
50	450	EU407803	0	517	EF417140	0	*435*	*U13929*	*0*
51	451	DQ408631	$5e^{-166}$	238	EF990722	$7e^{-115}$	*420*	*U73452*	*0*
52	461	DQ408631	$2e^{-165}$	530	EF990722	0	*411*	*U73452*	*0*
53	460	DQ408631	$1e^{-167}$	389	EF990722	0	*410*	*U73452*	*0*
54	666	X54253	0	566	EF417146	0	*459*	*EF514920*	*0*
57	442	U57030	$3e^{-156}$	571	EU195963	0	445	EU196006	0
58	626	EU652745	0	569	EU195963	0	437	EU196006	0
59	478	U57030	$3e^{-178}$	559	EU195963	0	492	AB051045	0
60				534	EU195963	0			
61				572	EU195963	0			
62				481	EF417140	0	*433*	*U13929*	*0*
63				532	EF990722	0			
64	665	FJ483518	0	639	EF417146	0	*494*	*EF514920*	*0*
65	591	AY830431	0						
66									
67				598	AY589395	0	*275*	*AF519229*	*$1e^{-134}$*
68									
69				218	AY589395	$2e^{-95}$			
70	581	AF519212	0	507	AY589395	0	489	AY589304	
71	284	AF519212	$6e^{-134}$						
72	345	AF519212	$4e^{-167}$	659	AY589395	0	490	AY589304	
73				571	EU195963	0			
74	568	U57030	0	567	EU195963	0	431	EU196006	0
75				467	AY589395	0			
76				496	AY589395	0			
77									
78				713	AY589395	0			
79				116	AM399047	$3e^{-24}$			
80	358	AF519212	$8e^{-151}$	360	AY589395	0			
81	426	U57030	$7e^{-158}$	568	EU195963	0	498	BR000309	0
82				569	EU195963	0			
83	448	U57030	$2e^{-164}$	576	EU195963	0	443	EU196006	0
84	454	U57030	$4e^{-161}$	565	EU195963	0	435	EU196006	0
85				608	DQ145655	0			
86				609	DQ145655	0			
87				756	AY589395	0			
88									
89				553	EF990725	0			
90				573	X03680	0			
91	598	U57030	0	571	EU195963	0	433	EU196006	0
92	409	U57030	$1e^{-148}$	569	EU195963	0	439	EU196006	0
93	415	DQ408632	$5e^{-141}$	570	EU195963	0	427	EU196006	0
94				573	EU195963	0			
95				748	AY589395	0			
96	175	AF519212	$5e^{-71}$						
Mean	524			557			423		

Chapter 4

DNA Barcodes and Video Capture and Editing (VCE): Integrating morphological vouchering with molecular diversity surveys.

4.1 Introduction

As DNA barcode projects become widespread, standard protocols have been developed so data from different projects can be integrated (such as ABBI[8] and TOL[9]). A barcode sequence is generated from a target gene from a specimen which usually has been identified to species. This is then used as an exemplar for the species that other sequences (and specimens) are compared to.

These projects mainly target charismatic megafauna, animals over 1 cm in length. These organisms are relatively easy to identify, as defining morphological characters are large and easily observed. There tends to be a lot of interest in these animals from a range of disciplines, such as behaviourists and ecologists as well as taxonomists. As human activities increase extinction rates, these animals tend to be at the centre of conservation campaigns and stimulate public interest. If megafaunal species disappeared, they would be missed, as they would no longer be seen. Obtaining DNA from type specimens of larger animals for barcoding is relatively simple and quickly done, without compromising the integrity of the DNA or the morphology. An individual is rarely destroyed in the process and can be re-sampled, for example another feather could be collected. When specimens are rare, can potentially go extinct, or re-sampling would prove difficult, then greater effort would be invested in the collection, documentation and preservation of a sample. Barcoding, and its applications, for megafaunal taxa tend to be straightforward to execute and fit within standard frameworks and definitions.

[8] All Birds Barcoding Initiative, http://www.barcodingbirds.org

[9] Tree of Life, USA, http://tolweb.org/tree/

Meiofauna, organisms with a body axis less than 2 mm, present different challenges for standardized barcoding. A lack of taxonomic expertise and species definitions inhibits species identification so specimens are unlikely to have been identified to species. At a cursory inspection, meiofauna can appear to be morphologically conserved, e.g. nematodes. However, there is morphological variation which is difficult to visualize using light microscopy. Scanning (SEM) and transmission electron microscopy (TEM) reveal true morphological diversity, hidden from the naked eye (De Ley and Bert, 2002). Species identification by SEM or TEM drastically increases the timescale and budget of a project, and becomes prohibitively expensive for large-scale surveys and excludes the possibility of DNA recovery (De Ley *et al.*, 2005). This means diversity surveys tend to be limited to genus level and the vast majority of meiofaunal specimens will not have a concrete species diagnosis or type specimens, especially from large-scale environmental surveys (De Ley *et al.*, 2005). So for meiofaunal barcoding, we have to accept that we might not have type specimens or type sequences.

In meiofaunal barcoding surveys, the process of obtaining and identifying specimens often means DNA is degraded beyond the minimum amount and length required. Moreover, the whole organism is digested to release DNA so maintaining a paratype specimen is not possible. Meiofaunal organisms are, by definition, small and such do not have large amounts of DNA. Moreover, specimens can be temperature sensitive such that leaving a sample at room temperature for a few hours will result in most specimens dying and rendering the DNA unusable for PCR. It is therefore critical to preserve DNA as quickly as possible to (preferably from live specimens) prevent the DNA breakdown by enzymes.

In order to preserve morphology and DNA, samples (e.g. soil extractions) can be split into two where one sub-sample would be preserved sympathetically for morphology and the other for DNA. Morphological preservation methods (i.e. formalin) destroy DNA, and some DNA preservatives do not maintain morphology. There is also the potential for discrepancies between the sub-samples for rare taxa. In order to document morphological diversity as well as sequence variation, a new preservative method is needed.

A promising solution is DESS. A solution of dimethylsuphoxide, EDTA, saturated with sodium chloride, which preserves morphology and DNA (Yoder *et al.*, 2006). Initially used as a preservative for avian blood (Seutin *et al.*, 1991), DESS also proved to be suitable for the preservation of morphology for up to six months (Dawson *et al.*, 1998). However, the preservative properties of DESS on either the morphology or DNA of meiofaunal organisms has not been extensively tested.

In addition to using DESS, integrating morphological vouchering into meiofaunal surveys could allow identification posthumously. It is possible to preserve type megafaunal specimens as museum accessions. Meiofaunal taxa can be preserved permanently fixed on slides, but this prohibits DNA collection. These physical objects can be damaged and lost and can only exist in a single location at any one time. Making a digital voucher of a specimen, whether a video or still image, can make these issues obsolete. Documenting a large animal by photographs or video is relatively easy. For microscopic organisms, this is a little more challenging, as the specimen needs to mounted and magnified. The most efficient method, in terms of isolating and recording a single specimen, is to create a temporary slide. Previously this temporary slide was documented by taking digital photographs stepwise through a specimen forming an image stack. However, even taking photographs with the smallest possible distance between focal planes through a nematode, some detail will be lost. Using a digital format also means that even at high resolutions, some information will be lost between pixels. When the images are stacked, the final file is often very costly in terms of computer memory.

As recording technology and equipment has advanced, the next step was to take multifocal video images instead of a series of stills. A video is recorded of the slide, whilst adjusting the focus by hand creating a virtual slide. By taking a video, structures can be followed through the body. Moreover, using high definition tape prevents the loss of detail than can be associated with using a digital, pixel based format. By using equipment and software which is publicly available, the raw video can be processed and compressed to smaller, manageable files (De Ley and Bert, 2002). These can be distributed

across the web or as hardcopies and made available to anyone. In this way, morphological information is not confined to single physical location.

A virtual slide has major advantages over a conventional collection such that the specimens cannot deteriorate or be damaged. Furthermore, they cannot be lost. There are issues associated with the protection and maintenance of digital information such as storage, curation and costs. These issues are being addressed by organisations such as CBOL where metadata of specimens and projects are being stored and accessed electronically. However, generating morphological vouchers for large numbers of meiofaunal specimens is not yet standard practice and has only been tested on some nematode taxa (De Ley *et al.*, 2005).

The investigations in this chapter were designed to explore the properties of DESS as a DNA and morphological preservative, the use of morphological vouchering with species identification keys, the integration of vouchering with small-scale meiofaunal diversity surveys and the performance of vouchering on other meiofauna taxa such as tardigrades, copepods and mites. Digital vouchering, if found to be suitable, could provide a method to document morphological diversity and posthumously identify taxa once the specimen had been destroyed for PCR (De Ley *et al.*, 2005). This would enable congruence, or discord, between morphological information and sequence data to be explored. For barcoding projects, COI has been suggested as a standard target (Hebert *et al.*, 2003b). However there are concerns regarding the universal nature of the COI primer sets available (particularly with nematode taxa (De Ley *et al.*, 2005), so in addition to COI, LSU and SSU were also used as barcode targets.

4.2 Methods

4.2.1 Testing DESS as a preservative

Melissa Yoder from Paul de Ley's lab at University California Riverside (UCR), California, USA) extracted nematodes from soil samples from UCR Department of Nematology, and preserved specimens in DESS in July 2005. Samples were shipped by mail to Edinburgh and stored at room temperature for nine weeks before individuals were picked for PCR. The samples were rinsed in sterile tap water for 30 – 60 minutes to remove salt residue and prevent inhibition of PCR. Individuals were picked, lysed in 20 μl 0.25 M NaOH and neutralised following Floyd *et al*. (2002). The resulting lysates were used as template DNA for SSU PCR in 20 μl total volume reactions using primers SSU_F_04 and SSU_R_26 (Table 4.1). PCR products were visualised on a 1.5 % agarose gel stained with 0.00002 % ethidium bromide with 1 kb DNA ladder (Invitrogen, Carlsbad, California, USA) as size markers. Positive products were cleaned following the Wellcome Trust Sanger Institute's protocols and sequenced at the University of Edinburgh Gene Pool sequencing service (http://www.genepool.ed.ac.uk) with DYEnamic ET Terminator chemistry on a 48-capillary ABI 3730 DNA Analyzer.

Sequence chromatograms were processed using trace2seq.pl (A. Anthony and M. Blaxter, unpublished). Sequences were compared to published sequences in GenBank (http://www.ncbi.nlm.nih.gov) by means of a BLASTn search (Altschul *et al*., 1997). Sequences which matched nematode sequences as a top BLAST hits were used in analysis.

To complement this experiment, nematodes were extracted from moss samples collected at the University of Edinburgh King's Building campus, Edinburgh, UK. Samples were sent by regular mail to UCR. Samples took 3-4 weeks to arrive. Individuals were picked for PCR and sequenced by Melissa Yoder at UCR. Specimens were also examined for retention of morphological features using light microscopy and scanning electron microscopy (SEM) by the UCR team.

This work contributed to Yoder *et al.* (2006), see Appendix 4.1.

4.2.2 Identifying known nematode species from Video Capture and Editing (VCE) clips

It was important to assess the compatibility of VCE clips with species keys used for identification. Several isolates of Aphelenchoidoidea species (class Chromadorea, order Tylenchida) were collected and multiple individuals were picked from several DESS preserved isolates. Dr Stephen Lewis, Clemson University, USA provided cultures of *Aphelenchus avenae* and *Aphelechoides fragariae* ^{SL}. Manuel Mota, University of Evora, Portugal supplied *Bursaphelenchus* cultures. Isolates of *Aphelenchoides subtenuis* and *Aphelenchoides fragariae* ^{CSL} were received from Central Science Laboratories, York, UK. Isolates were rinsed in deionised H_2O for 30 – 60 minutes to remove salt crystals which can obscure morphological detail. Multiple video clips were generated for each specimen at different magnifications to capture morphologically important regions (e.g. the head and tail) (see below). Individual specimens were then lysed in 20 µl micro-LysisPLUS (Microzone Ltd., West Sussex, UK) following the manufacturer's protocol. Lysates were used as template DNA for COI, LSU and SSU PCRs (Table 4.1). SSU PCRs were performed using SSU_F_04 and SSU_R_26.

PCR products were visualised on a 1.5 % agarose gel stained with 0.00002 % ethidium bromide with 1 Kbp DNA ladder (Invitrogen, Carlsbad, California, USA) as size markers. Positive products were cleaned following the Wellcome Trust Sanger Institute's protocols and sequenced at the University of Edinburgh Gene Pool sequencing service (http://www.genepool.ed.ac.uk) with BIG Dye Terminator chemistry on a 48-capillary ABI 3730 DNA Analyzer.

Sequence chromatograms were processed using trace2seq.pl (A. Anthony and M. Blaxter, unpublished). Nematode sequences were confirmed by means of a BLAST search (Altschul *et al.*, 1997) and top hit accession numbers and species from GenBank were recorded.

Identifications keys for Aphelenchoidoidea, developed by Dr Sue Hockland at Central Science Laboratories (CSL), York, and 'Aphelenchida, Longidoridae and Trichodoridae' (Hunt) were used to test the image quality and usability of the VCE clips to identify the isolate species.

VCE protocol

Image capture

A Sony Handycam HD 1080i (set to manual focus and recording high definition video (HDV) 1080i) and high definition (HD) videotapes were used to record multifocal images of specimens. If clips were recorded digitally, details could be lost as the image is stored as a set number of pixels. Although the number of pixels per image can be set high (e.g. 12 million per frame), the detail is not the same as when recorded on tape as the image is still condensed to fit the number of pixels. A unique VCE number was assigned to each individual so that all clips were traceable. This was recorded as a separate clip before the specimen and helped to identify between specimens during playback of the tape. Time-stamp and duration of clips were manually recorded as well as magnifications and specimen details (such as specimen number). Single specimens were mounted under a cover slip in a drop of ddH_2O on a modified fluorescent antibody slide (Gold Seal Products, Portsmouth, N.H., USA). This enabled the position of the specimen to be observed before the slide was imaged on a Zeiss Axiovert 135 microscope. The clip number was recorded before recording the entire specimen at low power (using either 10x or 20x objective depending on the size). Recordings started by focusing above the top of the specimen and recording as the fine focus of the microscope was used to focus through the body. This process was repeated at high power magnification using oil objectives to record different regions. The head and the tail regions were recorded for all specimens. If the nematode was an adult female, the vulva was also recorded. Any interesting morphological features were also recorded. Focusing through the specimen took up to 10 seconds to ensure all detail was recorded.

Image editing

Recordings were downloaded using Final Cut Pro 5.1 (Apple Inc., USA). New clips were created at each time break, creating separate clips for each region automatically. Clips were edited to 3 – 4 seconds using Compressor 2 version 2.3 (Apple Inc., USA) ensuring all relevant details were retained and

batch processed, creating QuickTime Pro files. Raw video was recorded at a frame size of 1920 x 1080 pixels (width by height). Using Compressor 2, frame size was reduced to a quarter of the original size (480x270 pixels) which is easier to view through the Internet. The video encoder used default settings for H.264 (for high-bandwidth connections) codec with deinterlacing algorithm set to blur and sharpen edge set to 5.0. Filenames were generated to include specimen number, magnification, and body region e.g. D12_10_1xpt55x63xtail.mov is the VCE clip for the tail region of specimen D12_10 at 1x .55x 63x magnification. Once full, original HD videos tapes were archived as well as downloaded and processed images.

The short QuickTime videos can then be released on the web. Videos can be played without stopping or the play pointer can be used to move up and down through the virtual slide.

Table 4.1 PCR details including primer names, sequences, expected length (number of bases), PCR conditions and references. All PCRs were performed using 2 μl of template, 2 μl 10x PCR buffer, 2 μl 0.2 mM dNTPs, 0.4 μl of 10 μM of each primer and 0.08 μl of *Taq* polymerase (Qiagen) in a 20 μl total reaction volume. For COI and LSU, the forward primers were used as sequencing primers. SSU_R_09 was used as the sequencing primer for all SSU PCR products (Blaxter *et al.*, 1998).

Target	Name	Sequence (5' - 3')	Length	PCR conditions	Ref
COI	LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G	658	94°C 60 sec; 5 cycles of 94°C 60 sec, 45°C 90 sec, 72°C 90 sec; 35 cycles of 94°C 60 sec, 50°C 90 sec, 72°C 60 sec; 72°C 300 sec	(Folmer *et al.*, 1994)
	HCO2198	TAA ACT TCA GGG TGA CCA AAA AAT CA			
LSU	D2A	ACA AGT ACC GTG AGG GAA AGT TG	≈890	95°C 300 sec; 5 cycles of 94°C 30 sec, 55°C 45 sec, 72°C 120 sec; 72°C 600 sec	(Ye *et al.*, 2007)
	D3B	TGC GAA GGA ACC AGC TAC TA			
SSU	SSU_F_04	GCT TGT CTC AAA GAT TAA GCC	893	94°C 300 sec; 35 cycles of 94°C 60 sec, 55°C 90 sec, 72°C 120 sec; 72°C 10 sec.	(Blaxter *et al.*, 1998)
	SSU_F_07	AAA GAT TAA GCC ATG CAT G			
	SSU_R_26	CAT TCT TGG CAA ATG CTT TCG			
	SSU_R_09	AGC TGG AAT TAC CGC GGC TG	565		

4.2.3 VCE integration into a meiofaunal diversity survey

Moss and soil samples were collected during the British Schools Exploring Society (BSES) Disko Island, Greenland Expedition from July to August 2007. The GPS co-ordinates were recorded and samples were air dried before being transported to Edinburgh University and stored at 4°C (Table 4.2). Samples were re-hydrated and extracted using a modified Baermann funnel approximately 8 - 10 weeks after collection. Extracts were preserved in DESS for approximately one year before VCE and barcoding was performed. Preserved samples were soaked in ddH_2O for up to one hour to dissolve salt crystals. Single specimens were processed using the VCE protocol, before PCR was performed. All specimens were recorded focusing from the top of the slide to the bottom. Whole body images were recorded at low magnification. Magnification was increased to video key morphological structures, such as the head and tail region nematodes or claws for tardigrades. Videos were edited and compressed following the VCE protocol (Section 4.2.2 above). Identifications were made after PCR and sequencing. The QuickTime files used to key out nematode specimens using 'De Nematoden van Nederland' (Bongers, 1994). Tardigrade specimens were putatively identified to genus using 'British Tardigrades' (Morgan and King), 'The Tardigrade Fauna of Greenland (Petersen) and 'Tardigrada of Southwest Virginia' (Riggin, 1962). Copepod and mite specimens identified using 'Microscopic life in Sphagnum' (Hingley).

After VCE, PCR was performed for three targets for all specimens, mitochondrial cytochrome oxidase c subunit I (COI), and nuclear small and large ribosomal subunit RNA (SSU and LSU respectively) (Table 4.1). SSU PCR was performed using the primer pair SSU_F_07 and SSU_R_26. PCR products were visualised on 1.5 % agarose gel stained with 0.00003 % SYBR Safe (Invitrogen, Eugene, Oregon, USA). Positive products were cleaned and sequenced using BIG Dye Terminator chemistry, as above.

Sequence chromatograms were processed using trace2seq.pl version 1.0.1 (A. Anthony and M. Blaxter, see http://www.nematodes.org/bioinformatics/trace2seq for download) and sequences were confirmed by using a BLASTn (Altschul *et al.*, 1997) search of GenBank.

MOTU_define.pl was used to cluster gene sequences into MOTUs using a range of cut-offs and multiple resamplings. Defined MOTUs were investigated for the stability of sequence membership to a particular MOTU and congruence to putative morphological identification from VCE clips. A sub-set of GenBank SSU nematode sequences (SSUdb) was created excluding unidentified or uncultured specimens. The SSUdb contained 2261 accessions where the specimen had been identified at least to family level. A BLASTn search was performed against the SSUdb of the SSU sequences to check putative morphological identification from the VCE clips.

4.2.4 Using VCE in a survey of tardigrades

Moss samples were collected by Melissa Yoder, UCR, USA, from Fort Lauderdale, Florida; Picnic Hill, UCR, California and Kõke'e State Park, Kaua'i, Hawaii in 2006 (Figure 4.1). Tardigrades were isolated from the moss extracts and preserved in DESS. During May and June 2009, individuals were picked, rinsed and prepared for VCE, lysed in 20 μl of micro-LysisPLUS and PCR performed for COI, LSU and SSU (using primers SSU_F_04 and SSU_R_26) (Table 4.1). PCR products were visualised on 1.5 % agarose gels stained with SYBR Safe (Invitrogen) following manufacturers protocols. Positive products were cleaned and sequenced using BIG Dye Terminator chemistry following Sanger protocols. Sequence chromatographs were processed using trace2seq.pl version 1.0.1 (A. Anthony and M. Blaxter, unpublished). Trimmed sequences were compared to GenBank entries by a BLASTn search (Altschul *et al.*, 1997) and non-tardigrade sequences removed from the dataset. The sequences from the six tardigrade specimens collected from Disko Island, Greenland (see section 4.2.3) were also included in data analysis. Sequences were clustered using MOTU_define.pl with a range of cut-off values and multiple resamples for the three gene sets. Minimum length of sequence and minimum overlap were set using the minimum sequence length of each gene set. Defined MOTUs were investigated for their robustness and concordance with morphological assignments from VCE clips. Putative morphological identifications were made using Morgan and King (1976).

Table 4.2 Disko Island samples collected in July 2007, including north and west GPS co-ordinates and the number of specimens picked.

Site	North	West	*n*	Comments
D01	69,34.846	053,31.174	0	Moss, animals dead
D02	69,34.795	053,31.161	0	Moss, animals dead
D04	69,34.868	053,31.195	24	Moss
D05	69,34.806	053,31.158	16	Moss
D06	69,34.782	053,31.136	32	Moss
D09	69,34.776	053,31.141	0	Soil sample, extract was very dirty, no specimens found
D11	69,34.449	053,31.161	29	Moss
D12	69,34.807	053,31.159	16	Moss
			117	

Figure 4.1 Map showing locations of tardigrade specimens. FD = Fort Lauderdale, Florida; CA = UCR, California; HI = Kōke'e State Park, Kaua'i, Hawaii; D = Disko Island, Greenland.

4.3 Results

4.3.1 DESS as a preservative

A total of 112 individuals were picked for SSU PCR, generating 81 positive PCR results. One sequence, JM_CAL_SSU_07, was less than 100 base pairs (bp) in length and was excluded from further analysis. Of the remaining 80 sequences, one was less than 300 bp (JM_CAL_SSU_61), eight sequences were between 434 and 499 bp long, and the remainder ranged from 500 to 526 bp. The mean sequence length was 506.3 bp (expected length of sequences was 565 bp).

A BLAST (Altschul *et al.*, 1997) search of GenBank revealed 57 specimens (71.3%) had a top hit to *Tylocephalus auriculatus* (Class Chromadorea, Order Araeolaimida, accession number AY284707) (blue cells, Table 4.3). Twelve specimens (15%) had top hits to *Acrobeloides thornei* (Class Chromadorea, Order Rhabditida, Family Cephalobidae, accession number EU543175). This included sequence JM_CAL_SSU_02. Although this sequence had a top hit to an uncultured tardigrade (accession number AM088448) (grey cell, Table 4.3). Further inspection revealed that of the top 50 BLAST hits, this was only tardigrade entry; the rest were nematode sequences and the top nematode hit was *A. thornei*, EU543175 and thus the tardigrade annotation is wrong. Seven sequences matched *Acrobeles complexus* (Family Cephalobidae, accession number AY284671) and two sequences matched *Steinernema glaseri* (Class Chromadorea, Order Rhabditida, Family Steinermatidae, AY284682). Two specimens matched sequence AJ966472, *Allodorylaimus* sp. which is in the Class Enoplea. All matches (except JM_CAL_SSU_61) had an E-value of 0.0.

DESS was also tested as a morphological preservative by Melissa Yoder at UCR (Yoder *et al.*, 2006). Initially, specimens appeared distorted after being placed in DESS but recovered their original appearance over a period of time. DESS preserved individuals prepared as glycerin mounts showed comparable morphological preservation to formalin preserved specimens (Yoder *et al.*, 2006).

Table 4.3 Results from nematodes picked from DESS preserved samples including the length of sequence, mean sequence length and top BLASTn hit from GenBank (accession number and species). The species listed for AM088448 was 'uncultured tardigrade'. However, the remaining 49 (of 50) top hits were nematode sequences. Sequence JM_CAL_SSU_61 (italics) had matched a nematode sequence and had a high E-value (2e $^{-138}$) so was included in the dataset.

Sequence name	Length bp	Top BLAST hit Accession	Species
JM_CAL_SSU_01	513	EU543175	*Acrobeloides thornei*
JM_CAL_SSU_02	510	AM088448	*Uncultured tardigrade**
JM_CAL_SSU_03	525	AY284707	*Tylocephalus auriculatus*
JM_CAL_SSU_04	506	AY284707	*T. auriculatus*
JM_CAL_SSU_05	525	EU543175	*A. thornei*
JM_CAL_SSU_06	505	AY284707	*T. auriculatus*
JM_CAL_SSU_08	504	AY284707	*T. auriculatus*
JM_CAL_SSU_09	525	EU543175	*A. thornei*
JM_CAL_SSU_10	525	EU543175	*A. thornei*
JM_CAL_SSU_11	510	AY284707	*T. auriculatus*
JM_CAL_SSU_12	507	AY284707	*T. auriculatus*
JM_CAL_SSU_13	506	AY284707	*T. auriculatus*
JM_CAL_SSU_14	525	AY284707	*T. auriculatus*
JM_CAL_SSU_15	506	AY284707	*T. auriculatus*
JM_CAL_SSU_16	524	EU543175	*A. thornei*
JM_CAL_SSU_17	526	AY284707	*T. auriculatus*
JM_CAL_SSU_18	507	AY284682	*Steinernema glaseri*
JM_CAL_SSU_19	514	AY284707	*T. auriculatus*
JM_CAL_SSU_20	438	AY284671	*Acrobeles complexus*
JM_CAL_SSU_21	501	AY284707	*T. auriculatus*
JM_CAL_SSU_22	508	AY284707	*T. auriculatus*
JM_CAL_SSU_23	525	AY284707	*T. auriculatus*
JM_CAL_SSU_24	524	AY284707	*T. auriculatus*
JM_CAL_SSU_25	522	EU543175	*A. thornei*
JM_CAL_SSU_26	470	AF202155	*T. auriculatus*
JM_CAL_SSU_27	526	AY284707	*T. auriculatus*
JM_CAL_SSU_28	526	AY284707	*T. auriculatus*
JM_CAL_SSU_29	500	AY284707	*T. auriculatus*
JM_CAL_SSU_30	506	AY284707	*T. auriculatus*
JM_CAL_SSU_31	516	EU543175	*A. thornei*
JM_CAL_SSU_32	505	AY284707	*T. auriculatus*
JM_CAL_SSU_33	511	AY284707	*T. auriculatus*
JM_CAL_SSU_34	434	AY284671	*Acrobeles complexus*
JM_CAL_SSU_35	523	AY284707	*T. auriculatus*
JM_CAL_SSU_36	523	AY284707	*T. auriculatus*
JM_CAL_SSU_37	523	AY284707	*T. auriculatus*
JM_CAL_SSU_38	487	AY284671	*Acrobeles complexus*
JM_CAL_SSU_39	486	AY284707	*T. auriculatus*

JM_CAL_SSU_40	505	AY284707	*T. auriculatus*
JM_CAL_SSU_41	526	AY284707	*T. auriculatus*
JM_CAL_SSU_42	448	AY284671	*Acrobeles complexus*
JM_CAL_SSU_43	523	AY284707	*T. auriculatus*
JM_CAL_SSU_44	506	AJ966472	*Allodorylaimus sp*
JM_CAL_SSU_45	507	AY284707	*T. auriculatus*
JM_CAL_SSU_46	524	AY284707	*T. auriculatus*
JM_CAL_SSU_47	507	AY284707	*T. auriculatus*
JM_CAL_SSU_48	506	AY284707	*T. auriculatus*
JM_CAL_SSU_49	516	AY284671	*Acrobeles complexus*
JM_CAL_SSU_50	511	AY284707	*T. auriculatus*
JM_CAL_SSU_51	507	AJ966472	*Allodorylaimus sp*
JM_CAL_SSU_52	523	EU543175	*A. thornei*
JM_CAL_SSU_53	524	AY284707	*T. auriculatus*
JM_CAL_SSU_54	501	AY284707	*T. auriculatus*
JM_CAL_SSU_55	506	AY284707	*T. auriculatus*
JM_CAL_SSU_56	484	AY284671	*Acrobeles complexus*
JM_CAL_SSU_57	505	AY284707	*T. auriculatus*
JM_CAL_SSU_58	506	EU543175	*A. thornei*
JM_CAL_SSU_59	506	AY284707	*T. auriculatus*
JM_CAL_SSU_60	504	AY284707	*T. auriculatus*
JM_CAL_SSU_61	282	AY284671	*Acrobeles complexus*
JM_CAL_SSU_62	506	AY284707	*T. auriculatus*
JM_CAL_SSU_63	523	AY284707	*T. auriculatus*
JM_CAL_SSU_64	524	AY284707	*T. auriculatus*
JM_CAL_SSU_65	524	AY284707	*T. auriculatus*
JM_CAL_SSU_66	506	AY284707	*T. auriculatus*
JM_CAL_SSU_67	517	AY284707	*T. auriculatus*
JM_CAL_SSU_68	506	AY284707	*T. auriculatus*
JM_CAL_SSU_69	525	AY284707	*T. auriculatus*
JM_CAL_SSU_70	523	EU543175	*A. thornei*
JM_CAL_SSU_71	506	AY284707	*T. auriculatus*
JM_CAL_SSU_72	521	AY284707	*T. auriculatus*
JM_CAL_SSU_73	506	AY284682	*Steinernema glaseri*
JM_CAL_SSU_74	505	AY284707	*T. auriculatus*
JM_CAL_SSU_75	522	AY284707	*T. auriculatus*
JM_CAL_SSU_76	507	EU543175	*A. thornei*
JM_CAL_SSU_77	503	AY284707	*T. auriculatus*
JM_CAL_SSU_78	506	AY284707	*T. auriculatus*
JM_CAL_SSU_79	499	AY284707	*T. auriculatus*
JM_CAL_SSU_80	505	AY284707	*T. auriculatus*
JM_CAL_SSU_81	524	AY284707	*T. auriculatus*
80	**506.3 bp**		

4.3.2 Identification of known species from VCE clips

Eight isolates from six species were used to investigate the usability of VCE clips to identify nematodes from keys. 93 individuals were picked, had VCE clips generated and PCR performed for multiple targets (Table 4.4). Sequences were checked by means of a BLAST search of GenBank to ensure the correct target (i.e. nematode and either COI, LSU or SSU) was amplified. No *Aphelenchus avenae* COI sequences were generated from these specimens. COI sequences have been generated from individuals, from the same isolate of *A. avenae,* which had not been processed through VCE. Moreover, there were only three LSU and two SSU sequences from the seven individuals. Primers for all three targets have been shown to work for *A. avenae* (unpublished data) and it is not clear why the success rate for these specimens is so poor.

Specimens from isolate *Aphelenchoides fragariae*^{SL} also failed to generate uncontaminated COI sequences. COI sequences from *Aphelenchoides subtenuis* and *Aphelenchoides fragariae* ^{CSL} specimens had top hits to *Ascaris suum* (Class Chromadorea, Order Ascaridida) and *Ancylostoma caninum* (Class Chromadorea, Order Rhabditida) respectively. There were no *Aphelenchoides subtenuis* COI sequences in GenBank so it would be impossible to get a top hit to this species. Although there were COI sequences from other members of the Family Aphelenchoididae (Order Tylenchida), the COI entry for *Aphelenchoides fragariae*, accession number AB067761, is not the same region amplified by the COI primers used in this study. The expectation would have been for the *Aphelenchoides* sequences to have top hits to members of Aphelenchoididae.

There is a single sequence from *Aphelenchoides subtenuis* in GenBank (accession number EF213109) which covers the 3' end of SSU, the internal transcribed spacer I (ITS I), 5.8S, ITS 2 and the 5' end of LSU. The regions amplified by the LSU and SSU primers in this study lie outside of this region and unsurprisingly the sequences generated did not BLAST match to this sequence. The LSU top hit was to *Schistonchus guangzhouensis* (DQ912927, Family Aphelenchoididae) and the SSU top hit was *Aphelenchoides blastophtorus* (AY284644). These results suggest that the specimens were from Aphelenchoididae, contrary to the COI results.

Interestingly, there were no positive LSU PCR results for *Bursaphelenchus luxuriosae* (Table 4.4). Both COI and SSU sequences were generated, indicating that there was DNA present. The LSU primers have been shown to work for a wide range of taxa and it is surprising that they did not work in this case. In this case, failure to generate positive PCR results (either for COI or LSU) does not seem to be caused by a lack of DNA in the lysate, as the other targets have been amplified. Sequences generated from *Bursaphelenchus conicaudatus* individuals had top hits to *B. conicaudatus* in GenBank for all three genes. Likewise, sequences from *Bursaphelenchus xylophilus* individuals had top hits to *B. xylophilus* sequences in GenBank for all targets.

VCE clips generated were 3 – 6 seconds in length and retained sufficient information to distinguish between species (Figure 4.2). Identification keys for the Aphelenchoididae are primarily based on adult morphology and, where present, male bursa. Specimens included males, females and juveniles, thus the key morphological structures used to distinguish among these species were tail shape and mucro, an extension of the cuticle at the tip of the tail (Figure 4.2, d - h). In most specimen clips, the mucro was clearly seen. Some specimens were distorted such that it was difficult to tell whether the head region was offset. In these specimens, the tail region was also distorted but it was still possible to distinguish the shape of the tail and mucro. The morphologies of the specimens investigated, broadly matched descriptions of Aphelenchoididae given by Dr. Sue Hockland (personal communication) and Hunt (1993). Although morphology was generally consistent, there were variations, notably the mucro of specimen 12 and phenotypic plasticity is known to occur (e.g. lateral lines) by Hunt (1993).

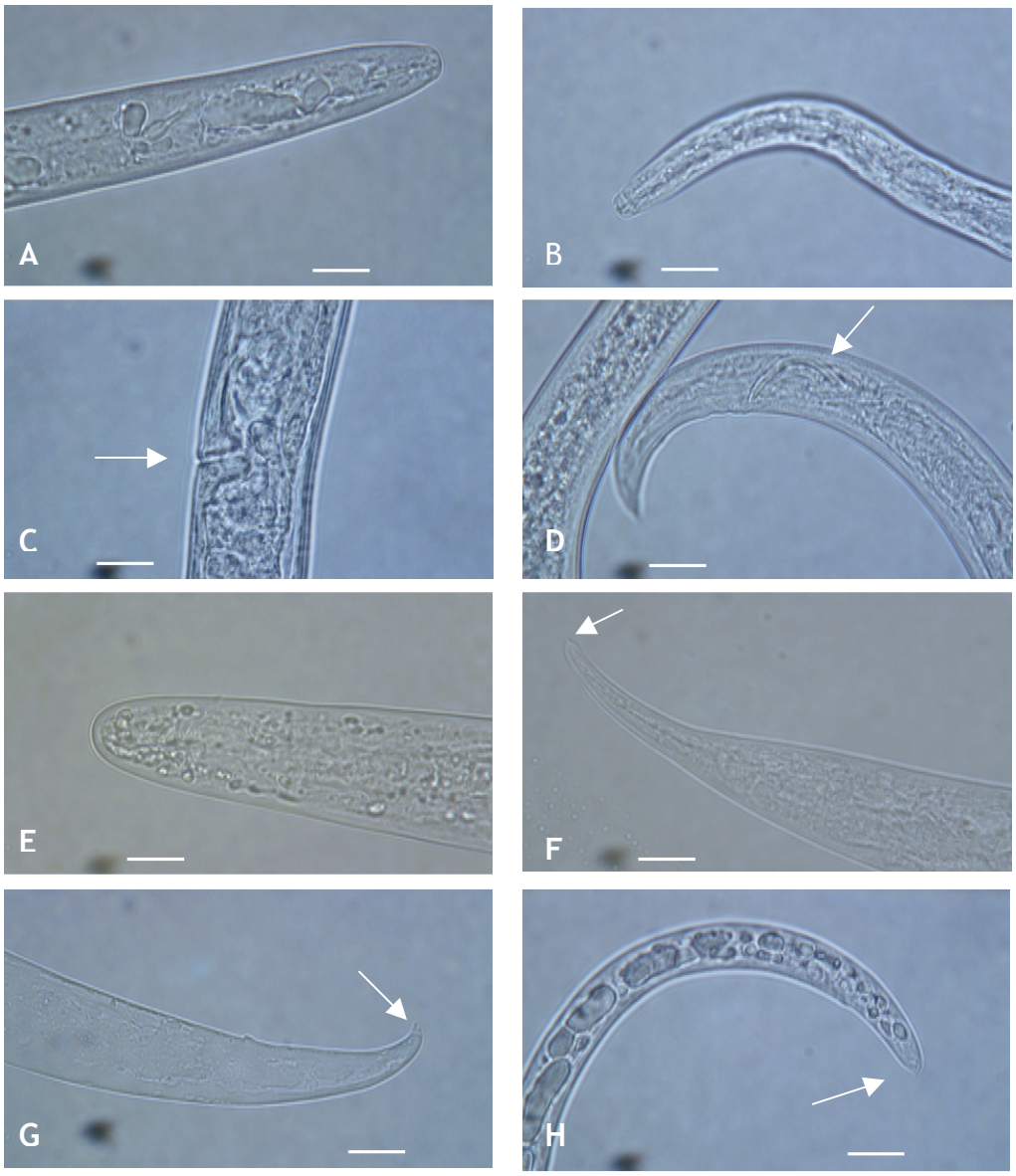
Table 4.4 List of specimens from identified species (A av = *Aphelenchus avenae*, B con = *Bursaphelenchus conicaudatus*, B lux = *Bursaphelenchus luxuriosae*, Aphd sub = *Aphelenchoides subtenuis*, B xyl = *Bursaphelenchus xylophilus*, Aphd fra = *Aphelenchoides fragariae*) used for VCE and to generate COI, LSU and SSU sequences. Head and tail shape including mucro morphology are described from VCE clips (Head shape: P/off = partial offset, E/con = elongated conoid. Tail shape: p = pointed, r = rounded, ^ indicates male specimens). See also Figure 4.2. [1] or [2] identify different isolates of the same species from the same source; different sources are indicated by [SL] (Stephen Lewis) or [CSL] (Central Science Laboratories. Mean sequence length and GenBank species top hits are listed. Shaded cells indicate failed sequencing of specimens, due to no sequence (-), short sequence (*) or non-nematode sequence generated. For BLAST hits, A av [A] accession number AB067763, A av [B] accession number AB368536, A. can (*Ancylostoma caninum*) accession number FJ483518, A. suum (*Ascaris suum*) accession number X54253, Aph sp [A] accession number AB368536, Aph sp [B] accession number AY284641, Aph sp [C] accession number AY284641, Aphd blas (*Aphelenchoides blastophtorus*) accession number AY284644, Aphd fra [A] accession number AB368540, Aphd fra [B] accession number AJ966475, Aphd sp [A] accession number FJ643488, Aphd sp [B] accession number FJ520227, Aphd sp [C] accession number FJ040407, Aphd sp [D] accession number FJ520227, Aphd sp [E] accession number DQ901552, B con accession number AB299227, B. frau (*Bursaphelenchus fraudulentus*) accession number AY508015, B lux accession number AB097863, B. xyl [A] accession number AY508070, B xyl [B] accession number AB067766, B xyl [C] accession number EU295504, B xyl [D] accession number EF446942, S. guang (*Schistonchus guangzhouensis*) accession number DQ912927, Uncultured (nematode) accession number AB385859, species codes as before.

ID #	Head shape	Tail shape	Mucro	Species	Top BLAST hit		
					COI	LSU	SSU
01	Blunt	Blunt	None	A av	-	A av [B]	-
02	Blunt	Blunt	None	A av	-	A av [B]	Uncultured
03	Blunt	Blunt	None	A av	-	A av [B]	Aph sp [B]
04	Blunt	Blunt	None	A av	-	-	-
05	Blunt	Blunt	None	A av	-	-	-
06	Blunt	Blunt	None	A av	-	-	-
07	Blunt	Blunt	None	A av	-	-	-
09	Offset	Conoid p	None	B con	A av [A]	B con	B. frau
10	Offset	E/con r	Single	B con	A av [A]	B con	B. frau
11	Offset	E/con r	Single	B con	A av [A]	B con	B. frau
12	Offset	E/con r	Trifurcate	B con	A av [A]	-	B. frau
13	Offset	Conoid p	Single	B con	A av [A]	-	B. frau
14	Offset	Conoid r	Single	B con	A av [A]	B con	B. frau
15	Offset	E/con r	Single	B con	A av [A]	B con	B. frau
16	Offset	E/con r	Single	B con	A av [A]	B con	B. frau
17	Offset	Conoid r	Single	B lux	-	-	-

18	Offset	Conoid r	Bifurcate	B lux	-	-	Aphd sp [B]
19	Offset	Conoid r	Single	B lux	B lux	-	Aphd sp [B]
20	Offset	Conoid p	Single	B lux	B lux	-	Aphd sp [B]
21	Offset	Conoid r	Single	B lux	B lux	-	Aphd sp [B]
22	Offset	Conoid r	Bifurcate	B lux	B lux	-	Aphd sp [B]
23	Offset	Conoid r	Bifurcate	B lux	B lux	-	Aphd sp [B]
24	Offset	Conoid r	Single	B lux	B lux	-	Aphd sp [B]
25	P/off	Conoid r^	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
26	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
27	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd sp [C]
28	P/off	Blunt	None	Aphd sub[1]	A. suum	S. guang	Aphd blas
29	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
30	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
31	P/off	Conoid r^	None	Aphd sub[1]	A. suum	S. guang	Aphd blas
32	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
33	P/off	Blunt	None	Aphd sub[1]	A. suum	S. guang	Aphd blas
34	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
35	P/off	Conoid r^	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
36	P/off	Conoid r^	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
37	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
38	P/off	Conoid r^	None	Aphd sub[1]	A. suum	S. guang	Aphd blas
39	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
40	P/off	Blunt	Single	Aphd sub[1]	A. suum	S. guang	Aphd blas
41	P/off	Conoid r^	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
42	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
43	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
44	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
45	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
46	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
47	P/off	Conoid r^	None	Aphd sub[2]	A. suum	S. guang	Aphd blas
48	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
49	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
50	P/off	Conoid r^	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
51	P/off	Conoid r^	None	Aphd sub[2]	A. suum	S. guang	Aphd blas
52	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
53	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
54	P/off	Conoid r^	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
55	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
56	P/off	Blunt	Single	Aphd sub[2]	A. suum	S. guang	Aphd blas
57	Offset	Conoid p^	None	B xyl[1]	B xyl [A]	B xyl [C]	Aphd sp [D]
58	Offset	Conoid p^	None	B xyl[1]	B xyl [A]	B xyl [C]	Aphd sp [D*]
59	Offset	Conoid p^	None	B xyl[1]	B xyl [B]	B xyl [C]	Aphd sp [D]
60	Offset	Conoid r	None	B xyl[1]	B xyl [A]	B xyl [C]	Aphd sp [D]
61	Offset	Conoid r	None	B xyl[1]	B xyl [B]	B xyl [C]	Aphd sp [D]
62	Offset	Conoid r	None	B xyl[1]	B xyl [B]	B xyl [C]	Aphd sp [D]
63	Offset	Conoid r	None	B xyl[1]	B xyl [B]	B xyl [C]	Aphd sp [D]
64	Offset	Conoid p^	None	B xyl[1]	B xyl [A]	B xyl [C]	Aphd sp [D]
65	Offset	Conoid p^	Single	B xyl[2]	B xyl [A]	B xyl [C]	Aphd sp [D]
66	Offset	Conoid r	None	B xyl[2]	B xyl [A]	B xyl [C]	Aphd sp [D]

67	Offset	E/con r	None	B xyl[2]	B xyl [A]	B xyl [C]	Aphd sp [D]
68	Offset	Conoid r	None	B xyl[2]	B xyl [B]	-	Aphd sp [D]
69	Offset	Conoid r	None	B xyl[2]	-	B xyl [D]	Aphd sp [D]
70	Offset	Conoid r	None	B xyl[2]	B xyl [A]	B xyl [C]	Aphd sp [D]
71	Offset	Conoid r	None	B xyl[2]	B xyl [A]	B xyl [C]	Aphd sp [D]
72	Offset	Conoid r	Single	B xyl[2]	B xyl [A]	B xyl [C]	Aphd sp [D]
73	P/off	Conoid r	Single	Aphd fra[CSL]	-	-	-
74	P/off	Conoid p	Single	Aphd fra[CSL]	*Chlorella*	-	*Koliella*
75	P/off	Conoid p	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
76	P/off	Conoid p	Single	Aphd fra[CSL]	Springtail	-	Aphd fra [B]
77	P/off	Conoid p^	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
78	P/off	Conoid p	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
79	P/off	Conoid p	Single	Aphd fra[CSL]	Springtail	Aphd fra [A]	Aphd fra [B]
80	P/off	Conoid p	Single	Aphd fra[CSL]	*Phytophtora*	Aphd fra [A]	Aphd fra [B]
81	P/off	Conoid p	Single	Aphd fra[CSL]	*Phytophtora*	Aphd fra [A]	Aphd fra [B]
82	P/off	Conoid p	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
83	P/off	Conoid p	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
84	P/off	Conoid p^	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
85	P/off	Conoid p	Single	Aphd fra[CSL]	-	Aphd fra [A]	Aphd fra [B]
86	P/off	Conoid p^	Single	Aphd fra[CSL]	*Phytophtora*	Aphd fra [A]	Aphd fra [B]
87	P/off	Conoid p	Single	Aphd fra[CSL]	*Phytophtora*	Aphd fra [A]	Aphd fra [B]
88	P/off	Conoid r	Single	Aphd fra[CSL]	*Phytophtora*	Aphd fra [A]	Aphd fra [B]
89	P/off	Conoid p^	Single	Aphd fra[SL]	A. can	Aphd sp [A]	Aphd sp [E]
90	P/off	Conoid r	Single	Aphd fra[SL]	A. can	Aph sp [A]	Aph sp [C]
91	P/off	Conoid r	Single	Aphd fra[SL]	-	Aph sp [A]	Aph sp [C]
92	P/off	Conoid p^	Single	Aphd fra[SL]	A. can	Aphd sp [A]	Aphd sp [E]
93	P/off	Conoid p^	Single	Aphd fra[SL]	A. can	Aphd sp [A]	Aphd sp [E]
94	P/off	Conoid p^	Single	Aphd fra[SL]	A. can	Aphd sp [A]	Aphd sp [E]
Total					**66** **586.9 bp**	**75** **652.5 bp**	**84** **460.3 bp**

Figure 4.2 Digital images from VCE clips of Aphelenchoidoidea specimens. A: Blunt head shape of specimen 02; B: Partially off-set head region of specimen 80; C: Vulval region (arrow) of specimen 91 (female); D: Bursa (arrow) of specimen 93 (male); E: Blunt tail of specimen 07; F: Elongated conoid shaped tail of specimen 15 with single mucro (arrow); G: rounded conoid tail of specimen 22 with bifurcate mucro (arrow); H: Blunt tail of specimen 45 with single mucro (arrow). (Scale bars: A – H = 100 μm).



4.3.3 Integration of VCE into a meiofaunal diversity survey

From the DESS preserved extracts, 117 specimens (six tardigrades, two mites, two copepods and 107 nematodes) were collected from five sites on Disko Island (Table 4.5). VCE clips were generated for all specimens before PCR was performed for COI, LSU and SSU.

Morphological identification of the non-nematode specimens was relatively simple to complete. Identification of the nematode specimens (Figure 4.3) was more time consuming and intricate due in part to the number of specimens. The majority of nematode specimens (67.2%) were identified as belonging to the class Chromadorea, with the family Plectidae having the most representatives. The remaining nematodes were from the class Enoplea, mainly from the order Dorylaimida (Table 4.5).

PCR was performed before specimens were morphologically identified. PCR success varied greatly between genes and taxa. There were 111 positive LSU PCRs which yielded 100 sequences once taxon identity had been confirmed by BLAST (85.4% specimens yielded correct sequences) (Appendix 4.2). There were 112 positive SSU PCRs products that were reduced to 105 sequences (89.7% success) (Appendix 4.2). Approximately a third of SSU sequences had top hits to GenBank entries which were identified as 'Uncultured tardigrade' (bold accession numbers, Table 4.5). When these GenBank entries were checked, they appeared to match nematode sequences, not tardigrade sequences. Thus, where a nematode specimen had a top hit to one of these 'uncultured tardigrades' it was counted as a nematode sequence. Of the 72 positive COI PCRs, only 39 sequences matched the morphological ID of the specimen (33.3% success) (Appendix 4.2) (Table 4.5). No sequences were generated for the two mite specimens (D06_28 and D12_13, pink cells Table 4.5). The LSU and SSU primers were originally designed for nematode taxa, but are known to work on a range of different taxa. Therefore theses primers might not be expected to have worked. COI primers LCO1490 and HCO2198 are supposedly universal, working across a broad range of taxa (Hebert *et al.*, 2003a) although other studies focussing on Arachinda have used different primers (Paquin and Hedin, 2004). It is possible that the lysis procedure failed for these mites. In comparison to other meiofaunal organisms, such as nematodes and tardigrades, mites have a thick cuticle

which may protect cells from the lysis buffer. This would result in none of the mite DNA being release into the lysate, but may cause material on the outside of the organism (such as algal cells) to lyse and release DNA.

All three targets were amplified from two copepod specimens (D11_03 and D11_27, green cells, Table 4.5). Five COI, four LSU and six SSU sequences were generated from the six tardigrade specimens (blue cells, Table 4.5). The LSU and SSU PCRs were successful for nematode specimens. In comparison, COI performed very poorly, (34.5% success). These results mirror the performance of the genes in Chapter 3 of this thesis and provide evidence against the suitability of COI as a universal barcode. Interestingly only one of the Tylenchida specimens had a positive PCR and generated a single SSU sequence (Table 4.5). This suggests that these primers are not optimised for this particular taxon of nematodes.

Percentages of BLASTn top hits to taxonomic levels were calculated for each of the gene datasets (Table 4.6). SSU sequences provided were the most specific taxonomic matches, 47.9% of the sequences had a top BLASTn hit to the same family as the putative morphological identification and 18.1% had top hits to the same genus (Table 4.6). The percentages of LSU matches were similar to the SSU results but the were more matches to the same order as the morphological identification. COI taxonomic matches were poor, with most specimens (87.2%) having a top hit to the same class as the morphological identification of the specimen (Table 4.6).

Figure 4.3 Selection of digital images from some of the Disko Island specimens. A: Copepod specimen D11_03; B: Tardigrade specimen D12_03 with eye spots (arrow); C; Rows of teeth (arrow) of Prionchulus specimen D04_03; D: Circular amphid (arrow) of Desmodoridae specimen D04_18; E: Head extensions and strong cuticular annulations of specimen D04_22; F: Head extensions and cuticular 'hairs' (arrow) of specimen D05_14; G: Typical Dorylaimida spear of specimen D06_02; H: Spiral amphid (arrow), 'hairs' and lip extensions of Achromadora specimen D11_05. (Scale bars: A – B = 100 μm; C – H = 10 μm).

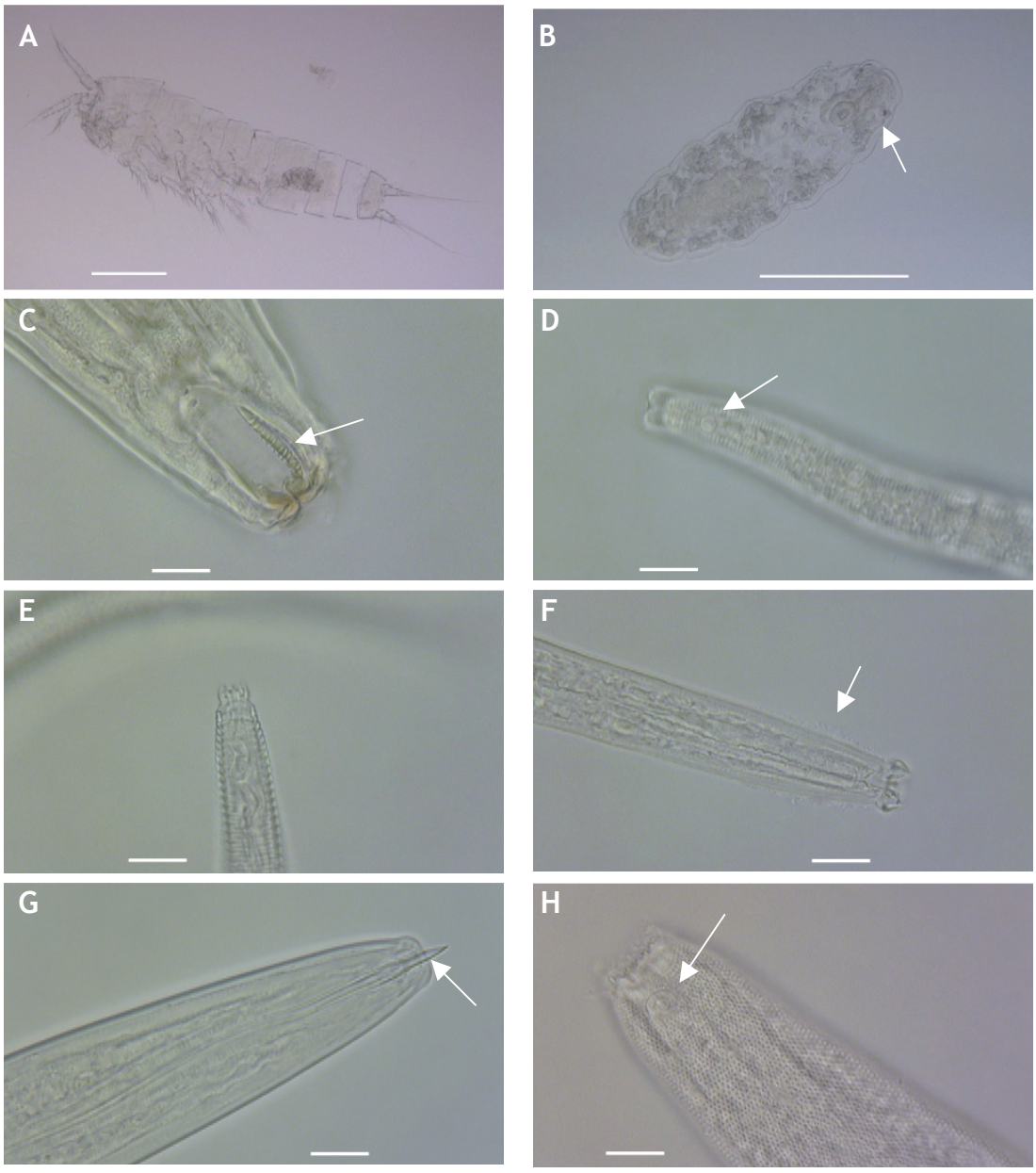


Table 4.5 List of meiofaunal specimens from Disko Island. Specimen ID includes location (see Table 4.2) and specimen number. Morphological identification to taxon level (P = Phylum, C = Class, O = Order, F = Family, G = Genus) is listed and the taxon level of match of the top hit is given for each gene. The level of taxonomic match to the top BLASTn hit for sequences are listed for COI, LSU and non-nematode SSU sequences. Accession numbers of the top hits for nematode SSU sequences are given. Total number of specimens, sequences and mean sequence length for each gene are also listed. Unsuccessful PCRs are indicated by – and – in grey cells indicate contaminant sequences. Specimen IDs in italics indicates specimens which failed to produce any usable sequences. Pink cells are mite specimens, green cells are the copepod specimens and blue cells are tardigrade specimens, the remainder are nematodes. GenBank accession numbers in bold for SSU sequences are identified as 'Uncultured tardigrade'. For nematode specimens, the level of match from the SSUdb to morphological ID is given (codes are the same as for COI and LSU BLASTn matches). ***** indicate specimens which were originally entered as *Eucephalobus* spp. but re-examination of VCE clips suggested specimens were morphologically similar to the other *Plectus* specimens in the samples.

Specimen ID	Morphological ID	Top Hit COI	LSU	SSU	SSUdb match
D04_01	Dorylaimida^{O}	-	O	EF024986	O
D04_02	Prionchulus^{G}	--	G	**AM088370**	G
D04_03	Prionchulus^{G}	--	G	**AM088370**	G
D04_04	Prodorylaimus^{G}	-	O	AY993978	O
D04_05	Dorylaimida^{O}	-	O	AJ966484	O
D04_06	Mesodorylaimus^{G}	-	O	AY993978	O
D04_07	Plectus^{G}	C	G	**AM088334**	F
D04_08	Prionchulus^{G}	--	G	AY284745	G
D04_09	Cephalobus^{G}	--	F	EU543175	F
D04_10*****	Plectus^{G}	C	F	**AM088334**	F
D04_11	Dorylaimida^{O}	-	O	AY993978	O
D04_12	Dorylaimida^{O}	-	O	AJ966484	O
D04_13	Plectus^{G}	-	-	**AM088354**	F
D04_14	Anaplectus^{G}	--	F	AY284696	G
D04_15	Cephalobus^{G}	--	F	AF430515	F
D04_16	Prionchulus^{G}	--	G	**AM088350**	G
D04_17	Dorylaimida^{O}	-	O	AY593946	O
D04_18	Desmodoridae^{F}	--	C	AY593935	F
D04_19	Cephalobus^{G}	--	F	EU543175	F
D04_20	Plectus^{G}	C	F	**AM088334**	F
D04_21	Dorylaimida^{O}	-	O	AY593951	O
D04_22	Teratocephalus^{G}	-	O	AY284683	G
D04_23	Plectus^{G}	-	G	AJ966513	F
D04_24	Prionchulus^{G}	--	G	**AM088350**	G
D05_01	Milnesium type	C	--	P	N/A

D05_02	Anaplectus^{G}	--	F	AY284696	G
D05_03	Plectus^{G}	C	G	**AM088334**	F
D05_04	Plectus^{G}	C	G	**AM088334**	F
D05_05	Plectus^{G}	C	C	AY284696	F
D05_06	Cephalobus^{G}	--	C	EU543175	F
D05_07	Dorylaimida^{O}	--	P	AY993978	O
D05_08	Plectus^{G}	--	--	**AM088334**	F
D05_09	Dorylaimida^{O}	-	--	AY593946	O
D05_10	Plectus^{G}	C	F	**AM088334**	F
D05_11	Plectus^{G}	-	F	**AM088342**	F
D05_12	Plectus^{G}	--	G	**AM088334**	F
D05_13	Achromadora^{G}	-	C	AY593941	G
D05_14	Metateratocephalobus^{G}	-	O	AY284686	G
D05_15	Plectus^{G}	C	P	**AM088354**	F
D05_16	Isohypsibius type	F	--	P	N/A
D06_01	Cephalobus^{G}	-	F	EU543175	F
D06_02	Dorylaimida^{O}	--	O	AY593946	O
D06_03	Tylenchida^{O}	-	O	--	N/A
D06_04	Tylenchida^{O}	-	--	--	N/A
D06_05	Cephalobus^{G}	--	F	EU543175	F
D06_06	Plectus^{G}	-	G	**AM088334**	F
D06_07	Tylenchida^{O}	-	--	--	N/A
D06_08	Plectus^{G}	C	F	AM088334	F
D06_09	Dorylaimida^{O}	-	O	EF024986	O
D06_10	Tylenchida^{O}	--	--	--	N/A
D06_11	Teratocephalus^{G}	-	O	AY284683	G
D06_12	Tylenchida^{O}	-	--	--	N/A
D06_13	Tylenchida^{O}	-	O	--	N/A
D06_14	Dorylaimida^{O}	-	O	EF024986	O
D06_15	Plectus^{G}	C	G	**AM088334**	F
D06_16	Tylenchida^{O}	-	--	--	N/A
D06_17	Plectus^{G}	--	F	AY284696	F
D06_18	Prionchulus^{G}	--	G	AY284745	G
D06_19	Dorylaimida^{O}	-	O	EF024986	O
D06_20	Plectus^{G}	-	G	**AM088334**	F
D06_21	Teratocephalus^{G}	-	O	AY284683	G
D06_22	Cephalobus^{G}	-	F	EU543175	F
D06_23	Pristionchus^{G}	C	--	--	N/A
D06_24	Dorylaimida^{O}	-	O	EF024986	O
D06_25	Dorylaimida^{O}	-	O	AY593951	O
D06_26	Dorylaimida^{O}	-	O	EF024986	O
D06_27	Dorylaimida^{O}	-	O	EF024986	O
D06_28	Mite	--	-	-	N/A
D06_29	Cephalobus^{G}	--	F	EU543175	F
D06_30	Dorylaimida^{O}	P	O	EF024986	O
D06_31	Tylenchida^{O}	-	-	-	-
D06_32	Cephalobus^{G}	-	F	EU543175	F
D11_01	Dorylaimida^{O}	--	O	AY993978	O
D11_02	Plectus^{G}	C	--	**AM088334**	F

D11_03	Bryocamptus type	C	C	G	N/A
D11_04	Plectus[G]	C	G	**AM088334**	F
D11_05	Achromadora[G]	-	O	AY593941	G
D11_06	Plectus[G]	C	F	**AM088334**	F
D11_07	Plectus[G]	C	F	**AM088334**	F
D11_08	Plectus[G]	C	F	**AM088334**	F
D11_09	Achromadora[G]	--	O	AY593941	G
D11_10	Plectus[G]	C	F	AJ966508	F
D11_11	Dorylaimida[O]	--	O	AJ966484	O
D11_12	Plectus[G]	C	F	**AM088412**	F
D11_13	Plectus[G]	C	F	**AM088354**	F
D11_14	Plectus[G]	C	F	**AM088412**	F
D11_15	Plectus[G]	C	F	**AM088412**	F
D11_16	Dorylaimida[O]	--	O	AY593946	O
D11_17	Anaplectus[G]	--	F	AY284696	G
D11_18	Plectus[G]	C	F	AJ966508	F
D11_19	Plectus[G]	C	F	**AM088412**	F
D11_20	Plectus[G]	C	F	**AM088412**	F
D11_21	Plectus[G]	C	F	**AM088412**	F
D11_22	Plectus[G]	C	F	**AM088412**	F
D11_23	Dorylaimida[O]	--	O	AJ966484	O
D11_24	Dorylaimida[O]	--	O	AY993978	O
D11_25	Plectus[G]	C	F	**AM088334**	F
D11_26	Plectus[G]	C	F	**AM088412**	F
D11_27	Bryocamptus type	C	C	G	N/A
D11_28	Anaplectus[G]	C	F	**AM088412**	F
D11_29	Dorylaimida[O]	--	O	AJ966484	O
D12_01	Plectus[G]	C	F	AJ966508	F
D12_02	Dorylaimida[O]	--	O	AY993978	O
D12_03	Macrobiotus var I	G	F	P	N/A
D12_04	Macrobiotus var I	G	F	P	N/A
D12_05	Macrobiotus var II	--	F	P	N/A
D12_06	Macrobiotus var II	G	G	P	N/A
D12_07	Tylenchida[O]	-	-	-	-
D12_08	Plectus[G]	C	F	AM088351	F
D12_09	Tylenchida[O]	-	-	AB376945	C
D12_10	Cephalobus[G]	--	F	EU543175	F
D12_11	Plectus[G]	-	G	AM088354	F
D12_12	Dorylaimida[O]	-	O	AY593951	O
D12_13	Mite	-	-	-	N/A
D12_14	Dorylaimida[O]	-	O	AY593946	O
D12_15	Dorylaimida[O]	-	O	AY593946	O
D12_16	Teratocephalus[G]	-	O	AM088371	G
117		**39**	**100**	**105**	**97**
		620.1bp	**719.3bp**	**506.5bp**	

Table 4.6 Breakdown of percentage of BLASTn top hit matches to taxonomic levels for Disko Island specimens.

Target	COI	LSU	SSU
Level of match			
Genus	7.7	17.0	18.1
Family	2.6	39.0	47.6
Order	-	36.0	27.6
Class	87.2	6.0	1.0
Phylum	2.6	2.0	5.7
Number of sequences	**39**	**100**	**105**

For each gene set, MOTU_define.pl was run with 100 resamples and the range of cut-offs used varied as mean sequence lengths varied. Cut-offs used for the SSU dataset were 0-20, 25, 30, 35, 40 45, 50 and 55 bases, which was approximately equal to 10% of the mean sequence length (506.5 bp). Minimum sequence length to include in the MOTU_define.pl run was set to 360 bp (the shortest sequence length was 363 bp from specimen D12_09, Table 4.5) and minimum overlap was set to 210 bp. The COI cut-off range was extended to include 60 and 65 bp. Minimum sequence length and minimum overlap were set to 400 bp and 240 bp respectively. The LSU mean sequence length was 719.3 bp so cut-off range was extended to include 70 and 75 bp. Minimum sequence length and minimum overlap were set to 450 bp and 270 bp respectively. All sequences were included in MOTU_define.pl analyses. For each cut-off, the mean number of MOTUs defined and standard deviation was calculated from the resample data (Figure 4.4).

For COI, the mean number of MOTUs defined from the 39 sequences at 0% cut-off was 37.4 ± 0.8, and continued to decrease up to a cut-off 1.8% (equivalent to 11 bp). At this cut-off, the mean number of MOTUs was stable at 23 (standard deviation was 0) until 3.1% (19 bp) after which it dropped to 22 until a cut-off of 4% (Figure 4.4). The mean number of MOTUs dropped from 22 to 16.5 as a 10.5% cut-off was reached. There is another plateau in the mean number of MOTUs (18.9 MOTUs) defined between the cut-off values of 7.2 and 8.1%. Within the first COI plateau, between cut-offs 1.9 and 3.1% (12 - 19 bp), the membership of sequences to MOTUs was investigated (Appendix 4.3). The mean number of MOTUs defined was 23 and standard deviation was 0.00. Over this range of cut-offs, not only did the number of MOTUs remained constant, but also the membership of sequences to MOTUs: all the MOTUs defined were equivalent.

The mean number of LSU MOTUs defined from the 100 sequences at 0% was 70.3 ± 1.1. This dropped to 40.5 ± 0.9 by a cut-off of 0.6 % (4 bp) (Figure 4.4). This was the largest initial decrease seen in any of the gene sets. There were more LSU MOTUs defined than COI or SSU up to cut-off 2.8% where the number of LSU MOTUs dropped below COI MOTUs and approached SSU results. The mean number of LSU MOTUs did not show a plateau as large as that seen in the COI results. After the initial sharp decrease, the number of

MOTUs defined levelled off between the cut-offs 1.1 and 1.4% (8 - 10 bp), from 36.4±0.5 to 36.0 ±0.7. At these cut-offs, the majority of MOTUs were equivalent, supported in at least 97% of the resamples (Appendix 4.3). There were two MOTUs defined in the primary run at a cut-off of 1.3% (9bp) which were not strongly supported by the resamples. These two MOTUs were more often merged.

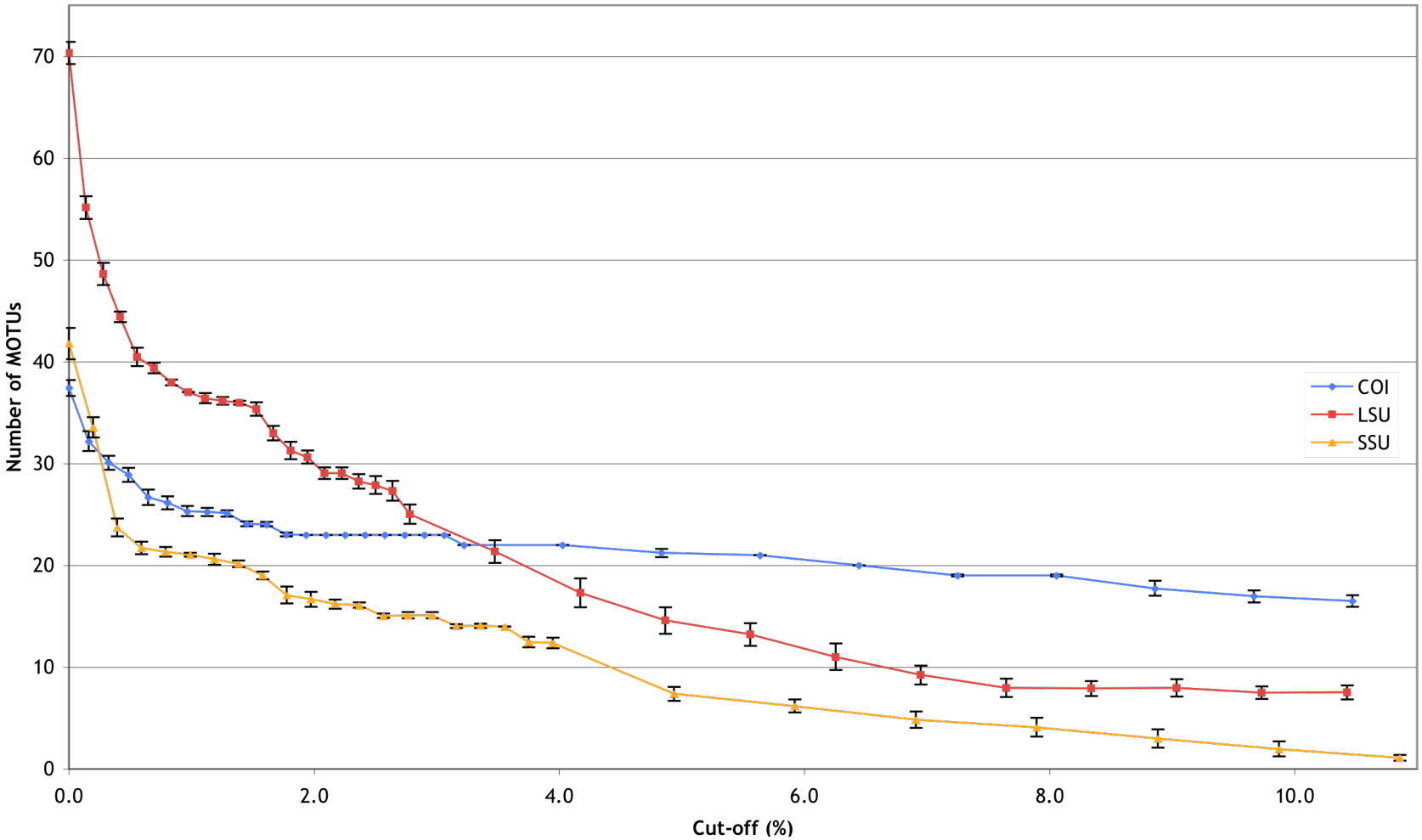
The mean number of SSU MOTUs defined from 105 sequences follows a similar pattern to COI and LSU results. There was initially a sharp decrease over the first few cut-offs (0 - 0.4%), after which the rate of decline decreased (Figure 4.4). There appeared to be several plateau phases in the SSU results. The first occurred between cut-offs 2.6 - 3.0% (13 – 14 bp) where the mean number of MOTUs defined was 15.0 - 15.1. There was another plateau between cut-offs 3.2 and 3.6% where the mean number of MOTUs defines was 14.0 (standard deviation ranged from 0.0 to 0.2). From the resamples, memberships of sequences to equivalent MOTUs were stable over both plateau phases (Appendix 4.3). The decrease from a mean of 15 MOTUs to 14 MOTUs was the result of two MOTUs merging at the higher cut-off values (3.2 - 3.6%). Interestingly, when the cut-off was increased to 9.9% (50 bp), the mean number of MOTU defined was 1.96 ± 0.75. At this cut-off there were nematode, tardigrade and copepod sequences clustered in the same MOTU. The most MOTUs were defined from the LSU dataset which was made up of 100 sequences. Interestingly, after the first two cut-offs, there were consistently more COI MOTUs defined than SSU MOTUs. The SSU dataset comprised 105 sequences, compared with 39 COI sequences. This suggests that there is less variation within the SSU sequences than the COI sequences. The morphological identifications of MOTUs were also investigated. In the COI, LSU and SSU plateau phases, copepod and tardigrade sequences formed separate MOTU to nematode sequences.

Morphological identification was investigated within the nematode MOTUs for the three genes. In the COI plateau phase, MOTUs were equivalent and stable at all the cut-offs and contained specimens from a single genus or family (Appendix 4.3). During the plateau phase in the LSU results, MOTUs were generally robust and equivalent. Although most LSU MOTUs contained sequences of specimens from the same family, some LSU MOTUs

(such as MOTU0014, MOTU0011 and MOTU0003 from the 9bp run) contained sequences from multiple families (Appendix 4.3). The SSU MOTUs were robust and equivalent across the plateau range. As in the COI results, MOTUs were restricted to a single family or genus (Appendix 4.3).

SSU nematode sequences were matched to the SSUdb by means of a BLASTn search. There were 2261 nematode sequences in the SSUdb, 20 of which were found to be top hits for the 97 nematode specimens. Most of the top hit of specimen sequences to the SSUdb were congruent with the putative morphological identifications made, at least to family if not genus (Table 4.5). Where morphological identification did not match SSUdb BLASTn results, the top five hits were examined. In all cases the top five hits belonged to the same genus or family. There was usually concordance between morphological and SSUdb identification at higher taxon levels (e.g. order) (Table 4.5).

Figure 4.4 Mean number of MOTUs defined at each cut-off value for all three genes and standard deviations from Disko Island specimens. Cut-off is expressed as a percentage of the mean sequence length for each gene.



4.3.4 VCE and tardigrade diversity

From the preserved samples, 98 specimens were picked, VCE clips generated and PCR performed following standard protocols. Disko Island tardigrade specimens were included in the analyse making a total of 104 specimens. VCE clips were used to define morpho-types, based on claw and median bulb structure. MOTU_define.pl was used to cluster the sequences.
From the VCE clips, four tardigrade genera were identified (Appendix 4.4). There was one *Echinischus* type specimen, HI14 (class Heterotardigrada, family Echiniscidae). Heterotardigrada are characterised by the presence of cuticle plates and four single claws on each leg (Morgan and King, 1976) (Figure 4.5 a and b). The rest of the specimens belonged to the class Eutardigrada. There were 19 specimens identified as *Milnesium tardigradum* (order Apochela, family Milnesiidae) which were easily identified by a pear-shaped pharyngeal bulb and unsymmetrical double claws with trifurcate secondary arms (Figure 4.5 c and d). Although some specimens appeared distorted from the DESS preservative (the internal structures had contracted and inverted the buccal region), it was possible to distinguish a variant type in which the papillae around the mouth were absent (Figure 4.5 e and f). There were 8 *Isohypsibius* type specimens (order Parachela, family Hypsibiidae) characterized by the two double unsymmetrical claws with elongated posterior branches (Morgan and King, 1976) (Figure 4.5 g).
Most of the specimens were *Macrobiotus* type (order Parachela, family Macrobiotidae) based on the claw structure of two double claws with symmetrical branches (Morgan and King, 1976) (Figure 4.5 h). Although claw structure was the primary feature within the *Macrobiotus* type specimens, variations were seen in pharyngeal bulb morphology and the presence or absence of macro- and microplacoids (Figure 4.5 i and j). There were also 5 specimens which had red pigmentation and almond-shaped pores visible on the cuticle (Figure 4.5 k), morphologically resembling *Macrobiotus occidentalis* (Morgan and King, 1976). These specimens were designated as Macrobiotus TIV (Table 4.7).
LSU and SSU PCRs were the most successful, generating 98 sequences from 102 specimens (96% success) each. There were 81 (79.4%) COI sequences generated. Two specimens failed to yield any sequences, FD02 and HI14

(Table 4.7). Specimen FD02 was morphologically very similar to FD01, FD03 and FD04 but had fewer macroplacoids (see Appendix 4.4). Specimen HI14 was the only *Echinischus*-type tardigrade found in the survey. This Class of tardigrades are characterized by the presence of a hard cuticle with plates which can make lysis difficult as the buffer is unable to penetrate the cuticle. It is probable that specimen HI14 did not lyse properly and failed to release DNA.

Twenty-three specimens failed to yield COI sequences. Fourteen of these specimens would have been expected to produce COI sequences as other specimens with the same LSU and SSU BLASTn results generated COI sequences (Table 4.7). Two out of four specimens, which failed to produce LSU sequences, would have been expected to generate LSU sequences. Four out of six failed SSU PCRs would have been expected to generate sequences (Table 4.7). Six Californian tardigrades (CA02, CA22, CA40, CA41, CA45 and CA47, yellow cells, Table 4.7) BLAST matched the same named species as the top hit for LSU and SSU sequences; *Hypsibius convergens*. All six failed to produce COI sequences. This would indicate that the COI primers do not work for this species. The 98 specimens do not include any other *Hypsibius* sp. based on the BLAST results, so it is not clear whether the failure of these six COI PCRs was a random failure or primer specificity issue (i.e. the primers do not work in this species or genus). Also, no COI sequences were obtained from the Florida (FD) specimens. The two SSU sequences generated from FD01 and FD03 had a top hit to AJ617459, which was found to match the BLAST results of some of the Hawaii sequences. This would suggest that it should have been possible to generate COI sequences from the Florida specimens. However, the BLAST results of the LSU sequences did not match any of the top hits from either the California or Hawaii BLAST results. The SSU results suggest that it should have been possible to generate COI sequences, however it is not possible to say whether the LSU results support this. Given that FD04 generated BLAST results that were the same as FD01 and FD03, a SSU sequence from FD04 would have been expected.

The sequences were clustered for each gene set using MOTU_define.pl. For the COI run, minimum sequence length was set to 480 bp and minimum overlap was 479 bp. For LSU analysis, minimum sequences length was 230

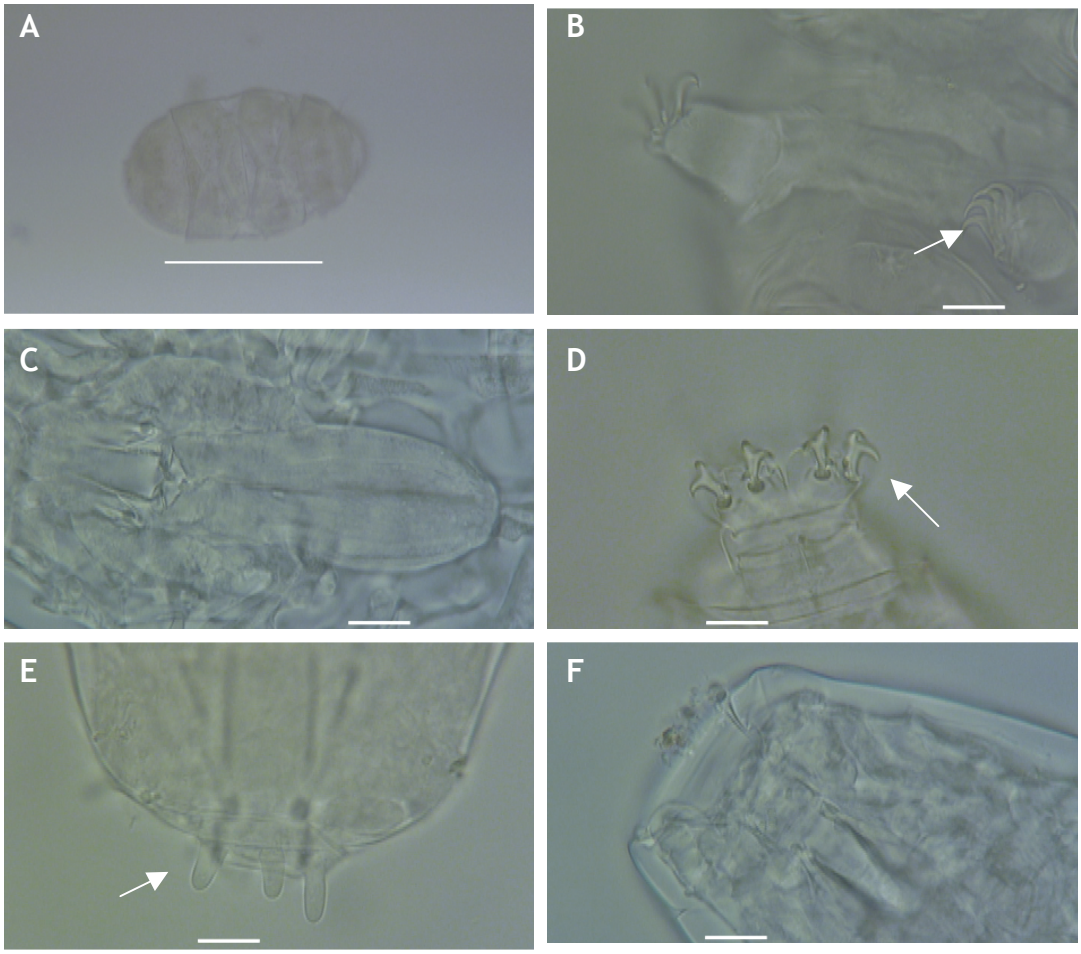
bp and minimum overlap was 229 bp. Minimum sequence length and overlap was set to 340 bp and 339 bp respectively for the SSU run. Clustering of sequences was investigated over the same range of cut-off values (0-25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80 bp). The resamples were used to calculate means and standard deviations of the number of MOTUs defined at each cut-off (Figure 4.6).

At 0 bp cut-off, there were 51.6 ±1.5 COI MOTUs defined from 81 sequences (Figure 4.6). As the cut-off was increased, there was a sharp decline in the mean number of MOTUs defined up to a cut-off of 1.4% (9 bp). The mean number of COI MOTUs defined plateaus from cut-offs 1.4 - 9.4% (9 - 60 bp) at 12 MOTUs. After a cut-off of 9.4%, the mean number of defined MOTUs decreases at a much slower rate than in the initial phase.

In the LSU data set there were 57.1 ±2.2 MOTUs defined at 0% cut-off from 98 sequences. There is a sharp initial decline, similar to that seen in the COI results (Figure 4.6), before the rate of decrease slows. There is also a plateau phase in the LSU results between cut-offs 2.2 - 3.1% (16 - 22 bp) where the mean number of MOTUs defined falls from 14.28 ±0.47 to 13.51 ±0.52. The mean number of MOTUs defined decreased after the cut-off 4.2% before levelling off again at 7.6%.

The mean number of SSU MOTUs defined followed a similar pattern to the COI and LSU results. An initially a sharp decline in the number defined levelled off in a plateau phase, followed by gentler decline. At 0% cut-off, there were 31.6 ±1.06 SSU MOTUs defined (Figure 4.6), fewer than the mean number of COI and LSU MOTUs defined at 0% cut-off. The plateau phase was between the cut-offs 1.38 - 2.95% (7 - 15 bp) where there were approximately 11 MOTUs defined.

Figure 4.5 Key morphological features used to segrated specimens into tardigrade genera and types. A: Low magnification of whole specimen HI14 showing cuticular plates; B: Third pair of legs and claws (arrow) from specimen HI14; C: Pharyngeal bulb of specimen HI30; D: Trifurcate secondary claws (arrow) on hind legs of specimen HI12; E: Papillae (arrow) present in buccal region on specimen HI18; F: Papillae absent from buccal region on specimen HI30; G: Unsymmetrical *Isohypsibius* type claws (arrow) on hind leg of specimen CA02; H: Symmetrical *Macrobiotus* type claws on hind leg of specimen FD02; I: Oval pharyngeal bulb with elongated macroplacoids (arrow) from specimen CA10; J: Circular pharyngeal bulb with macroplacoids from specimen HI38; K: Red pigmentation and almond-shape pores (arrows) from *Macrobiotus* TIV specimen HI10. (Scale bars: A = 100 μm; B – K = 10 μm).



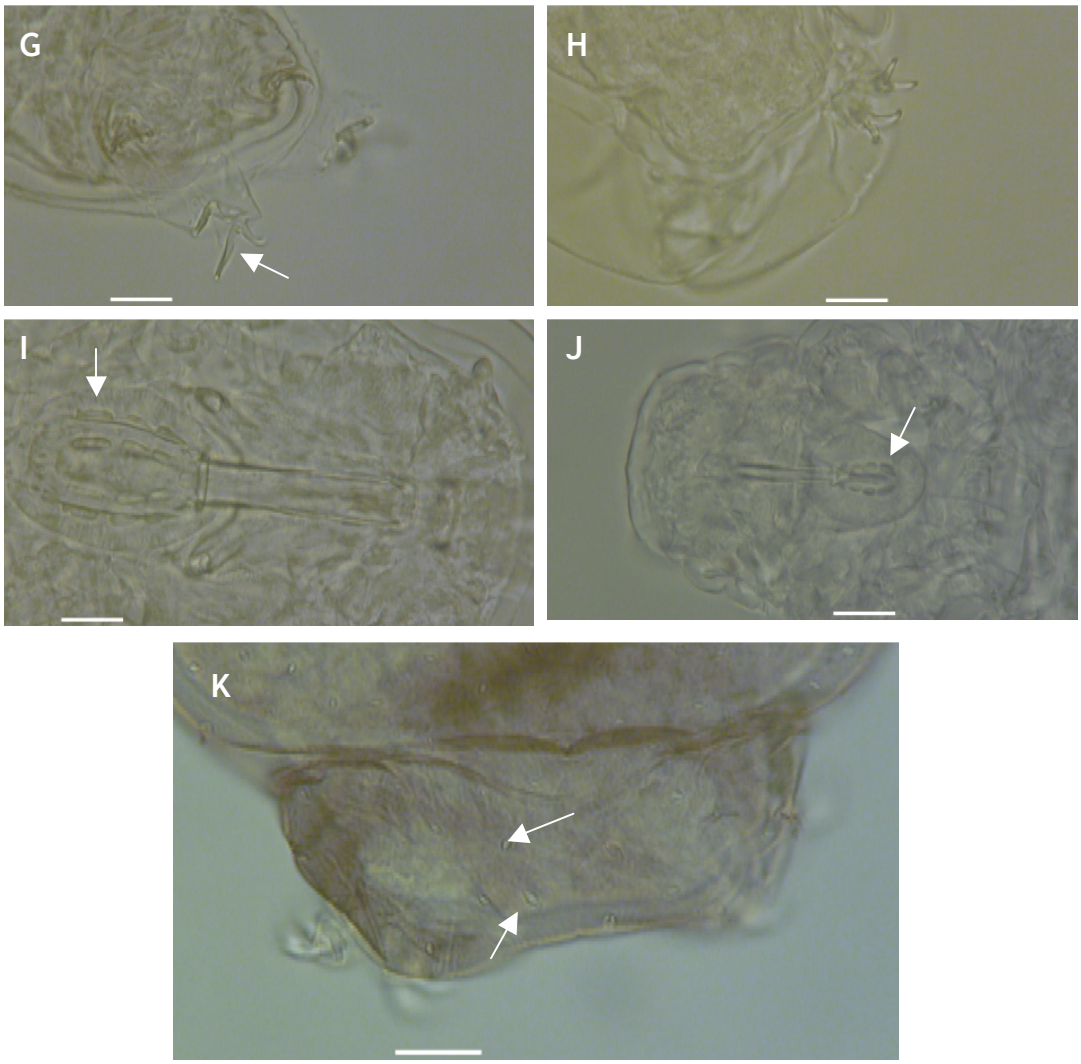

G
H
I
J
K

Table 4.7 Details of tardigrade specimens collected in 2006 from USA states and Greenland in 2007. Specimen ID includes the origin of the sample (FD = Florida; CA = California; HI = Hawaii; D = Disko Island) and specimen number. GenBank accession numbers and species of top hits are given. Where GenBank accessions are not a named species (i.e. AY210826 is described as *Milnesium* spp.), the first named species is listed as they were from the same genus. Total numbers of specimens and sequences generated (along with mean sequence length) for each gene are listed at the bottom of the table. Failed PCRs are indicated by -. PCRs which would have been expected to work, based on the results from the other genes, are indicated by ∗ (13 COI, 2 LSU and 4 SSU PCRs). Specimen names in italics indicate that no sequences were generated. ⅄ these accession numbers have 'Tardigrada sp.' as a species name, where Tardigrada relates to the Phylum, not a species. In all cases, the first named species to appear in the top hits was '*Hypsibius* cf *convergens*' (accession number AM500650). ✣ these accession numbers also have 'Tardigrada sp.' as a species name, but the top named species hit was *Macrobiotus sapiens* (accession number DQ839601). Accession numbers with ★ indicate 'Uncultured tardigrade' sequences where the top named species was *Macrobiotus richtersi* (accession number EU038081), † indicates top named species was *Macrobiotus tonollii* (accession number DQ839605), ✦ indicates top named species was *Ramazzottius oberhaeuseri* (accession number AY582122) and ♣ indicates top named species was *Diphascon* spp. (accession number EU266951). Genera codes are *A.* = *Astatumen*, *D.* = *Diphascon*, *H.* = *Hypsibius*, *I.* = *Isohypsibius*, *Ma.* = *Macrobiotus*, *Mil.* = *Milnesium*, *Min.* = *Minibiotus*, *Mu.* = *Murrayon*, *R.* = *Ramazzottius*.

Specimen ID	Morpho-type	Top Hit COI	Species	LSU	Species	SSU	Species
Florida							
FD01	Macrobiotus TI	- ∗		FJ435759	*Min. furcatus*	AJ617459✣	*Ma. sapiens*
FD02	Macrobiotus TI var I	-		-		-	
FD03	Macrobiotus TI	- ∗		FJ435759	*Min. furcatus*	AJ617459✣	*Ma. sapiens*
FD04	Macrobiotus TI	- ∗		FJ435759	*Min. furcatus*	- ∗	
California							
CA01	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA02	Isohypsibius TI	-		FJ435771	*H. convergens*	AJ617430⅄	*H.* cf *convergens*
CA03	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*

CA04	Macrobiotus TIII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA05	Macrobiotus TIII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA06	Macrobiotus TII var I	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA07	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088234★	*Ma. richtersi*
CA08	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	- ∗	
CA09	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA10	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA11	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA12	Macrobiotus TIII var I	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA13	Macrobiotus TII	- ∗		FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA14	Milnesium TI	- ∗		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
CA15	Macrobiotus TII var II	FJ435807	*Ma. pallarii*	- ∗		- ∗	
CA16	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA17	Macrobiotus TIII var I	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088202★	*Ma. richtersi*
CA18	Macrobiotus TII var II	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA19	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA20	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088208★	*Ma. richtersi*
CA21	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA22	Isohypsibius TI var I	-		FJ435771	*H. convergens*	AJ617430⅄	*H.* cf *convergens*
CA23	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088264★	*Ma. richtersi*
CA24	Macrobiotus TV	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA25	Macrobiotus TII var I	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA26	Macrobiotus TIII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088234★	*Ma. richtersi*
CA27	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088206★	*Ma. richtersi*
CA28	Macrobiotus TII var V	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA29	Macrobiotus TIII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA31	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA32	Isohypsibius TI var II	FJ176212	*Ma. macrocalix*	DQ077800	*Isohypsibius* spp	EF620403	*I. granulifer*
CA33	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	- ∗	

CA34	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA35	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA36	Macrobiotus TII var V	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA37	Macrobiotus TII var II	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA38	Macrobiotus TIII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA39	Macrobiotus TII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA40	Isohypsibius TI var I	-		FJ435771	*H. convergens*	AJ617430⅄	*H.* cf *convergens*
CA41	Isohypsibius TI	-		FJ435771	*H. convergens*	AJ617430⅄	*H.* cf *convergens*
CA42	Macrobiotus TV	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA43	Macrobiotus TII var I	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088118★	*Ma. richtersi*
CA44	Macrobiotus TIII	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA45	Isohypsibius TI var I	-		FJ435771	*H. convergens*	AJ617451⅄	*H.* cf *convergens*
CA46	Macrobiotus TII var V	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088138★	*Ma. richtersi*
CA47	Isohypsibius TI var I	-		FJ435771	*H. convergens*	AJ617430⅄	*H.* cf *convergens*
CA48	Macrobiotus TII var V	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088252★	*Ma. richtersi*
CA49	Macrobiotus TIII var I	FJ435807	*Ma. pallarii*	FJ435756	*Ma. pallarii*	AM088264★	*Ma. richtersi*
Hawaii							
HI01	Macrobiotus TI	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI02	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI03	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI04	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI05	Macrobiotus TIII	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI06	Macrobiotus TIII var I	-		FJ435761	*Min. gumersindoi*	DQ839602	*Ma. areolatus*
HI07	Milnesium TI	- *		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI08	Macrobiotus TI ?	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI09	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI10	Macrobiotus TIV	FJ176217	*Ma. macrocalix*	FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI11	Milnesium TI var I	- *		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI12	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI13	Macrobiotus TV	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*

HI14	Echinischus TI	-		-		-	
HI15	Milnesium TI var I	- *		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI16	Milnesium TI	- *		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI17	Macrobiotus TI var I	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI18	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI19	Macrobiotus TI var I	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI20	Macrobiotus TV	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI21	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI22	Macrobiotus TIV	FJ176217	*Ma. macrocalix*	FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI23	Macrobiotus TIV	FJ176217	*Ma. macrocalix*	FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI24	Macrobiotus TII	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI25	Macrobiotus TV var I	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI26	Macrobiotus TIV	FJ176217	*Ma. macrocalix*	FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI27	Macrobiotus TIII var I	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI28	Isohypsibius TI	EU244598	*Ma.*cf *richtersi*	DQ077800	*Isohypsibius* spp	EF620404	*I. prosostomus*
HI29	Macrobiotus TV	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI30	Milnesium TI var I	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI31	Macrobiotus TI	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI32	Macrobiotus TII var IV	- *		FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI33	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI34	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI35	Milnesium TI	EU244604	*Mil. tardigradum*	AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI36	Macrobiotus TII var III	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI37	Milnesium TI var I	- *		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*
HI38	Isohypsibius TI	EU251380	*R.* cf *oberhaeuseri*	- *		EF620404	*I. prosostomus*
HI39	Macrobiotus TII var III	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI40	Macrobiotus TII var III	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI41	Macrobiotus TIV var I	FJ176212	*Ma. macrocalix*	FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI42	Milnesium TI var I	- *		AY210826	*Mil. tardigradum*	U49909	*Mil. tardigradum*

HI43	Macrobiotus TII var III	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI44	Macrobiotus TII var III	EU251382	*R.* cf *oberhaeuseri*	FJ435755	*Ma. hufelandi*	DQ839601	*Ma. sapiens*
HI45	Macrobiotus TII var III	FJ435801	*Mu. dianeae*	FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
HI46	Macrobiotus TII	- ∗		FJ435762	*Mu. dianeae*	AJ617459✣	*Ma. sapiens*
Disko							
D05_01	Milnesium TII	FJ435798	*H. convergens*	AY593049		AJ617461✦	*R. oberhaeuseri*
D05_16	Isohypsibius TII	FJ435792	*A. trinacriae*	EF417148		AM088135♣	*Diphascon* spp.
D12_03	Macrobiotus TV var II	EU244598	*Ma. richtersi*	FJ435761	*Min. gumersindoi*	AM088251 †	*Ma. tonollii*
D12_04	Macrobiotus TV var II	EU244609	*Ma. tonollii*	FJ435761	*Min. gumersindoi*	AM088207 †	*Ma. tonollii*
D12_05	Macrobiotus TV var III	- ∗		FJ435759	*Min. furcatus*	AJ617459✣	*Ma. sapiens*
D12_06	Macrobiotus TV var III	FJ176207	*Ma. macrocalix*	FJ435755	*Ma. hufelandi*	AM088437✣	*Ma. sapiens*
104		**81** **636.2 bp**		**98** **719.7 bp**		**98** **508.1 bp**	

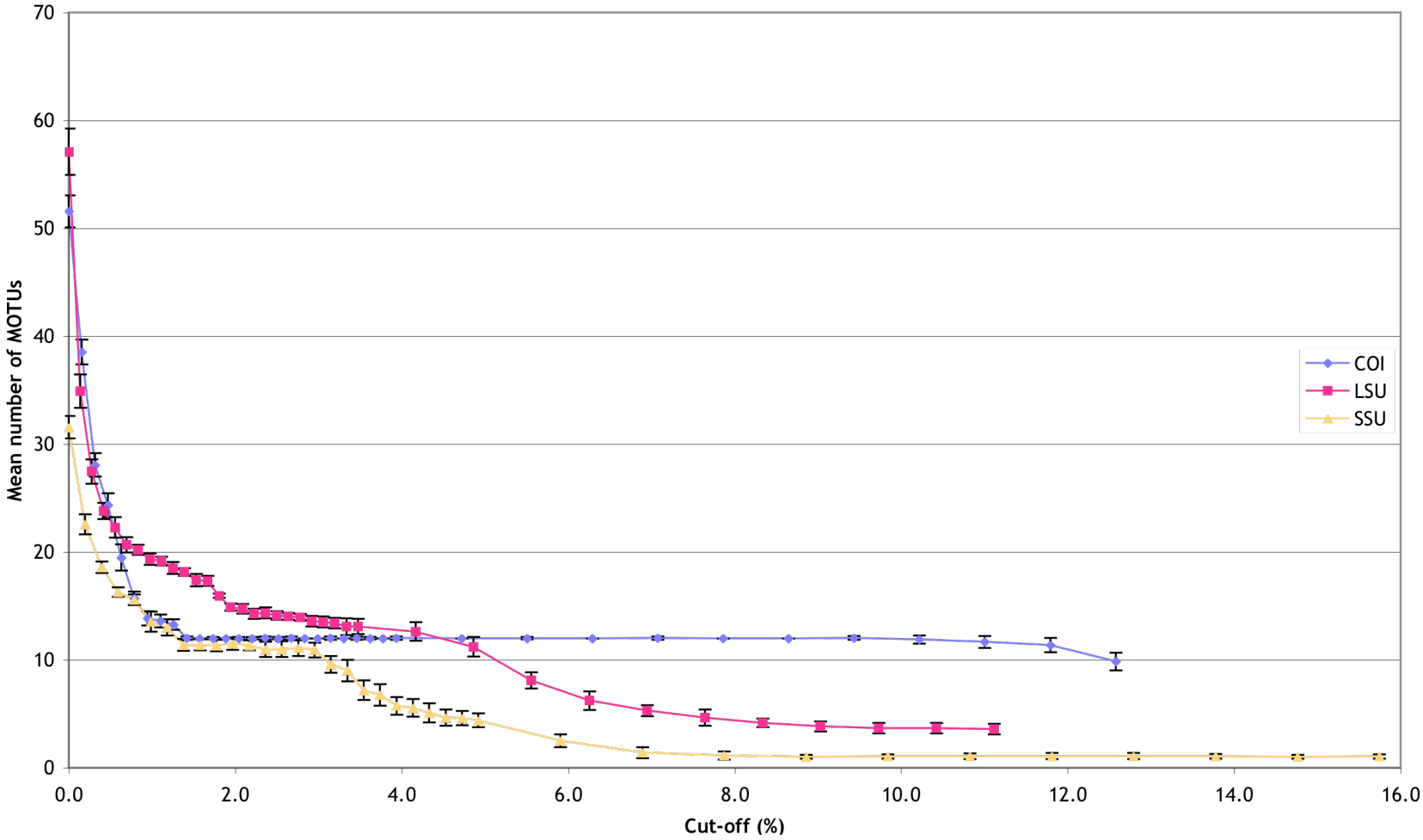
Within the COI plateau phase of 1.4 – 9.4% (9 – 60 bp), memberships of MOTU were stable. The re-sample data showed that MOTUs equivalent to those defined in the primary run were found in at least 97 out of 100 re-samples (Appendix 4.5). Moreover, different MOTUs contained sequences from only one of the sampling site (i.e. Hawaii, California, Disko Island or Florida).

The least amount of change in the mean numbers of LSU MOTUs defined was seen between 2.2 and 2.8% (16 - 20 bp). The memberships of sequences to MOTUs were investigated over this cut-off range. Most LSU MOTUs were stable, with equivalent MOTUs found in at least 90 of the re-samples. The more robust MOTUs tended to have fewer member sequences (Appendix 4.5). The robust MOTUs contained specimens from only one sampling site.

In the plateau phase of the SSU results, the membership of MOTUs was broadly stable at the lower cut-offs. However the stability decreased in some MOTUs as the upper cut-off value for the plateau phase was reached (Appendix 4.5). Some SSU MOTUs were made up of sequences from a single site, but unlike the COI MOTUs, others contained sequences from multiple sites.

Morphological identifications of specimens were also examined across the plateau phases of each gene. Within in the COI plateau phase, MOTUs were broadly congruent with morphological designations. Specimens identified as *Isohypsibius* formed MOTUs with no other specimens (Appendix 4.5). The same was found for *Milnesium* type specimens and those identified as *Macrobiotus* TIV. Specimens identified as *Macrobiotus* TI, TII, TIII and TV formed mixed MOTUs. Within the LSU plateau phase, not all MOTUs were robustly supported, thus MOTUs usually contained a mixture of morphological types (e.g. *Milnesium* and *Macrobiotus* TI, TII, TIII and TV) (Appendix 4.5). However, MOTUs containing only *Isohypsibius* or *Macrobiotus* TIV specimens were strongly supported. These MOTU were also supported by in the SSU results in addition to *Milnesium* specimens. As in the COI and LSU MOTUs, specimens identified as *Macrobiotus* TI, TII, TIII and TV cluster together in the same MOTUs (Appendix 4.5).

Figure 4.6 The mean number of MOTUs defined at each cut-off for all three genes of tardigrade specimens, with standard deviations. Cut-offs are expressed as a percentage of the mean sequence length.



4.4 Discussion

4.4.1 DESS is a preservative

Barcode surveys of meiofaunal organisms have been a compromise between preserving morphological detail for the traditional taxonomists and maintaining DNA integrity for barcoding. Morphological preservation techniques are not able to preserved DNA and vice versa. Splitting a sample to preserve both morphology and DNA is not ideal as rare taxa, represented by few individuals may be in one sub-sample but not the other. Moreover, traditional preserving methods for morphology and DNA use hazardous materials (such as formalin and ethanol respectively), can require special conditions and are problematic to ship using postal or courier services (Yoder *et al.*, 2006).

The results from this investigation have shown that DESS is a good DNA preservative. It is easy to prepare, samples can be stored at ambient temperature and transported without specialist equipment or precautions. These qualities make DESS an ideal preservative not only in the lab, where freezer space can be limited, but also for fieldwork. Moreover DNA can be amplified from specimens that have been stored for up to three years (see section 4.2.3, this chapter).

In addition, DESS preserves morphology. Although the high salt concentration of the solution can distort morphology, particularly of organisms with soft cuticles, rinsing with ddH_2O allows recovery of body shape. This step also removes salt crystals which may obscure internal morphological detail and is necessary for successful PCR.

DESS works well as a morphological and a DNA preservative is cheap and safe.

4.4.2 Utility of VCE for identifying known species

VCE clips can retain sufficient morphological detail to identify some nematode specimens to genus level and can identify laboratory cultures. In environmental surveys, VCE clips can distinguish between nematode orders,

at worst, and at best identify genera. Some specimens can be distorted due to preservation methods which could hinder identification. A greater obstacle to species identification, especially for meiofaunal taxa, is the lack of concrete species descriptions and definitions. Initially, VCE is unlikely to cement species descriptions but will highlight the true extent of morphological variation, and when used in conjunction with barcoding will provide a molecular framework for species definitions.

VCE should prove very effective for documenting morphological variation in meiofaunal surveys where 'type' specimens are likely to be rare as the whole animal is destroyed in the process of barcoding. It also circumvents issues with slide preservation. Online databases of VCE clips that are linked to sequence information could provide major assets for meiofaunal taxonomy. They can provide a permanent, mobile record of type specimens. For new taxa, it could provide a record of type features that would enable direct comparisons with identified taxa. It could also provide a record of known morphological variations within a species which would prevent multiple descriptions for a single species being recorded and taxonomic conflicts. This would also encourage communication between international taxonomists. Finally, it could provide users with a method of comparing unknown taxa. NemATOL (http:// nematol.unh.edu) currently provides an interactive database containing molecular, morphological, ecological and phylogenetic information for the nematode community, but there are other phyla which constitute meiofauna. It should be a simple process to replicate the NemATOL format and expand it to cover other phyla such as Tardigrada and Rotifera.

The number of unclassified taxa has been estimated to be several millions. With an increase in extinction rates, finding a system that can integrate molecular and morphological data would help to measure the range of diversity on Earth. VCE vouchering could provide a quick, cheap method which can be easily integrated into current barcoding protocols which may help to catalogue life on Earth.

The VCE method does require practice. There were significant improvements, both in efficiency and quality of specimens and images, the more specimens that were processed. The surveys described in this chapter,

were carried out solely by the author. Although there is a standard VCE protocol, differences between investigators could produce slight variations between clips. In larger scale surveys where multiple investigators produce clips, care should be taken to ensure consistency.

4.4.3 Integration of VCE into meiofaunal surveys

VCE has been demonstrated to be sufficient for documenting morphological variation of nematodes (De Ley *et al.*, 2005; Yoder *et al.*, 2006). When used as part of a barcoding survey, it can provide a putative identification which can be linked with molecular data. Key to integrating VCE into a barcode survey is using DESS, preserving both morphology and DNA so both can be obtained from a single specimen. Initially developed for nematode taxa, this investigation has demonstrated that VCE is easily applicable to other meiofaunal taxa. Any organism which can be made into a temporary slide should be suitable for VCE. Tardigrades often contract when put into DESS and it can be difficult to plainly see the claws and pharyngeal bulb. Making the temporary slide causes tardigrades to be flattened under the coverslip, thus allowing structures to be seen more clearly. Single specimen PCRs from environmental surveys are quick to perform: the specimen is picked from the substrate or extract and transferred to an individual tube before being lysed. If the VCE protocol is followed, a specimen is picked, temporarily slide mounted and clips generated before lysis and PCR. Although this does take longer, with practice it is possible to process 32 specimens in a day. If the sample has been preserved in DESS, then there is no rush to process everything. The quickest environmental barcode surveys use a bulk extract from the substrate and a single PCR reaction which generates sequences from multiple taxa. There is no way to record morphology using a bulk extraction method as everything is collected and then pooled. In addition, setting up a VCE system within a laboratory is quick and simple.
It is also possible to check initial morphological designation once molecular results have been generated. In the meiofaunal survey of Disko Island, specimens D04_10 and D05_11 were originally recorded as Cephalobidae

specimens. However molecular results indicated they were similar to other *Plectus* specimens collected in the survey. Re-examination of the VCE clips confirmed specimens D04_10 and D05_11 were morphologically most similar to D04_07 and D05_10 respectively. The specimens did not resemble other Cephalobidae specimens and the most probable cause for the incorrect identification being recorded was an error during data entering.

There appear to be mis-matches between morphological identification and some MOTUs in the tardigrade survey (Appendix 4.6). This may be due to the process of identifying specimens from VCE clips and over-estimating the amount of morphological variation within the specimens. VCE clips were easily sufficient to differentiate between genera of tardigrade specimens and these were supported by defined MOTUs. However, the features used to differentiate among Macrobiotus TI, TII, TIII and TV do not seem to relate to molecular differences.

There was an interesting result from the Disko Island meiofaunal survey. In the SSU MOTU analysis, where as the cut-off approached 10%, the mean number of MOTU defined approached one. As barcoding surveys are normally targeted to specific taxa, this would not be an issue. However, in this survey specimens derived from multiple phyla. At a cut-off of 9.9%, approximately two SSU MOTU were defined which included specimens from Nematoda, Tardigrada and Arthropoda. If this result is found in other meiofaunal surveys, then it may be necessary to re-assess the utility of the SSU gene to act as a barcode marker, depending on the aims of the survey.

Appendix 4.1

Appendix 4.2

List of VCE numbers and BLASTn top hit matches for Disko Island specimens. Symbols and colours as for Table 4.5.

Specimen ID	VCE number	Top Hit COI	LSU	SSU
D04_01	199	-	AY593010	EF024986
D04_02	200	CP000849	DQ077802	AM088370
D04_03	201	AY500368	DQ077802	AM088370
D04_04	202	-	AY592999	AY993978
D04_05	203	-	AY593036	AJ966484
D04_06	204	-	AY592998	AY993978
D04_07	205	EU407802	AY652779	AM088334
D04_08	206	CP000683	DQ077802	AY284745
D04_09	207	AY737269	DQ077788	EU543175
D04_10	208	X54253	AY652779	AM088334
D04_11	209	-	AY592998	AY993978
D04_12	210	-	AY593036	AJ966484
D04_13	211	-	-	AM088354
D04_14	212	EF159692	EF417148	AY284696
D04_15	213	EU768917	DQ077788	AF430515
D04_16	214	CP000409	DQ077802	AM088350
D04_17	215	-	AY593034	AY593946
D04_18	216	EU266375	DQ086700	AY593935
D04_19	217	AY737269	DQ077788	EU543175
D04_20	218	AJ558163	AY652779	AM088334
D04_21	219	-	AY593020	AY593951
D04_22	220	-	EF990727	AY284683
D04_23	221	-	EF417148	AJ966513
D04_24	222	CP000849	DQ077802	AM088350
D05_01	223	FJ435798	AY593049	AJ617461
D05_02	224	AF120660	AY652779	AY284696
D05_03	225	EU652745	EF417148	AM088334
D05_04	226	AJ558163	EF417148	AM088334
D05_05	227	AF200830	DQ077754	AY284696
D05_06	228	AY737269	AY821763	EU543175
D05_07	229	EF650564	AY652779	AY993978
D05_08	230	CP000409	DQ077800	AM088334
D05_09	231	-	DQ077800	AY593946
D05_10	232	AJ558163	AY652779	AM088334
D05_11	234	-	AY652779	AM088342
D05_12	233	CP000766	EF417148	AM088334
D05_13	235	-	AY652779	AY593941
D05_14	236	-	DQ077788	AY284686
D05_15	237	EU407799	AY592998	AM088354
D05_16	238	FJ435792	EF417148	AM088135
D06_01	239	-	DQ903085	EU543175

D06_02	240	--	AY593039	AY593946
D06_03	241	-	AY780971	Z70526
D06_04	242	-	DQ079800	Z70526
D06_05	243	AY737269	DQ077788	EU543175
D06_06	244	-	EF417147	AM088334
D06_07	245	-	--	Z70526
D06_08	246	EU407799	AY652779	AM088334
D06_09	247	-	AY593020	EF024986
D06_10	248?	-	DQ079800	Z70526
D06_11	250	-	EF990727	AY284683
D06_12	249	-	DQ079800	Z70526
D06_13	251	-	AY780971	Z70526
D06_14	252	-	AY593010	EF024986
D06_15	253	EU407780	AY652779	AM088334
D06_16	254	-	DQ079800	Z70526
D06_17	255	EF989696	AY652779	AY284696
D06_18	256	EF473798	DQ077802	AY284745
D06_19	257	-	AY593020	EF024986
D06_20	258	-	EF417147	AM088334
D06_21	259	-	EF990727	AY284683
D06_22	260	-	DQ903085	EU543175
D06_23	261	U57030	DQ079800	Z70526
D06_24	262	-	AY593010	EF024986
D06_25	263	-	AY593020	AY593951
D06_26	264	-	AY593010	EF024986
D06_27	265	-	AY593020	EF024986
D06_28	266	AF462318	-	-
D06_29	267	AY737269	DQ077788	EU543175
D06_30	268	AY591323	AY593010	EF024986
D06_31	269	-	-	-
D06_32	270	-	DQ903085	EU543175
D12_01	271	EU407798	AY652779	AJ966508
D12_02	272	--	AY592999	AY993978
D12_03	273	EU244598	FJ435761	AM088251
D12_04	274	EU244609	FJ435761	AM088207
D12_05	275	--	FJ435759	AJ617459
D12_06	276	FJ176207	FJ435755	AM088437
D12_07	277	-	-	-
D12_08	278	EU407801	AY652779	AM088351
D12_09	279	-	-	AB376945
D12_10	280	AY737269	DQ077788	EU543175
D12_11	281	-	EF417148	AM088354
D12_12	282	-	AY593020	AY593951
D12_13	283	-	-	-
D12_14	284	-	AY593036	AY593946
D12_15	285	-	EF207240	AY593946
D12_16	286	-	EF990727	AM088371
D11_01	287	CP000849	AY592999	AY993978
D11_02	288	EF043402	AY210813	AM088334

D11_03	289	AF462318	AY210813	AY627015
D11_04	290	EU346694	AY652779	AM088334
D11_05	291	-	DQ077754	AY593941
D11_06	292	DQ408627	AY652779	AM088334
D11_07	293	EF043402	AY652779	AM088334
D11_08	294	EU407780	AY652779	AM088334
D11_09	295	EF057733	DQ077754	AY593941
D11_10	296	EU652745	AY652779	AJ966508
D11_11	297	DQ317045	AY593036	AJ966484
D11_12	298	EU407780	AY652779	AM088412
D11_13	299	EU407780	AY652779	AM088354
D11_14	300	EU407780	AY652779	AM088412
D11_15	301	EU407780	AY652779	AM088412
D11_16	302	CR628336	AY593036	AY593946
D11_17	303	--	AY652779	AY284696
D11_18	304	EU407780	AY652779	AJ966508
D11_19	305	EF043402	AY652779	AM088412
D11_20	306	DQ408627	AY652779	AM088412
D11_21	307	EU407780	AY652779	AM088412
D11_22	308	DQ408627	AY652779	AM088412
D11_23	309	AM749297	AY593036	AJ966484
D11_24	310	EF650542	AY592999	AY993978
D11_25	311	EU407780	AY652779	AM088334
D11_26	312	EU407780	AY652779	AM088412
D11_27	313	EF554835	AY210813	AY627015
D11_28	314	DQ408627	AY652779	AM088412
D11_29	315	CP000849	AY593036	AJ966484

Appendix 4.3

Resample details of Disko Island MOTUs from plateau phases compared MOTUs defined in the primary run. Number, member sequences and taxa (Ac = *Achromadora*; An = *Anaplectus*; C = Copepod; Des = Desmodoridae; Do = Dorylaimida; Mes = *Mesodorylaimus*; Met = *Metateratocephalus*; Pl = Plectidae; Prio = *Prionchulus*; Pris = *Pristionchus*; Pro = *Prodorylaimus*; T = Tardigrada; Ter = *Teratocephalus*; Ty = Tylenchida) of each MOTU are listed. Resample MOTUs are classed as E: Equivalent to primary run; S: Split in comparison to primary run; J: MOTUs joined in comparison to primary run; C: MOTUs are rearranged as complex (i.e. split and joined) in comparison to primary run.

COI	#	Member Sequences	Taxa	E	S	J	C	Total
12bp_MOTU23	1	D05_04	Pl	100	0	0	0	100
12bp_MOTU22	1	D12_06	T	100	0	0	0	100
12bp_MOTU21	1	D11_09	Ac	100	0	0	0	100
12bp_MOTU20	1	D04_14	An	100	0	0	0	100
12bp_MOTU19	2	D12_03 D12_04	T	100	0	0	0	100
12bp_MOTU18	1	D12_01	Pl	100	0	0	0	100
12bp_MOTU17	1	D05_02	An	100	0	0	0	100
12bp_MOTU16	1	D06_17	Pl	100	0	0	0	100
12bp_MOTU15	1	D05_16	T	100	0	0	0	100
12bp_MOTU14	2	D11_18 D06_15	Pl	100	0	0	0	100
12bp_MOTU13	1	D11_27	C	100	0	0	0	100
12bp_MOTU12	2	D04_07 D06_08	Pl	100	0	0	0	100
12bp_MOTU11	3	D05_10 D05_15 D04_20	Pl	100	0	0	0	100
12bp_MOTU10	1	D04_15	Ce	100	0	0	0	100
12bp_MOTU09	3	D11_02 D11_07 D11_19	Pl	100	0	0	0	100
12bp_MOTU08	1	D06_30	Do	100	0	0	0	100
12bp_MOTU07	1	D11_03	C	100	0	0	0	100
12bp_MOTU06	1	D05_05	Pl	100	0	0	0	100
12bp_MOTU05	1	D04_10	Pl	100	0	0	0	100
12bp_MOTU04	14	D11_22 D11_28 D11_04 D11_14 D11_08 D11_15 D11_13 D11_21 D11_10 D11_26 D11_20 D11_12 D11_25 D11_06	Pl, An	100	0	0	0	100
12bp_MOTU03	1	D06_23	Pris	100	0	0	0	100
12bp_MOTU02	1	D05_01	T	100	0	0	0	100
12bp_MOTU01	2	D05_03 D12_08	Pl	100	0	0	0	100
12bp_MOTU	23			2300	0	0	0	2300
13bp_MOTU23	1	D05_05	Pl	100	0	0	0	100
13bp_MOTU22	1	D11_09	Ac	100	0	0	0	100
13bp_MOTU21	1	D04_15	Ce	100	0	0	0	100
13bp_MOTU20	1	D11_03	C	100	0	0	0	100
13bp_MOTU19	1	D05_04	Pl	100	0	0	0	100
13bp_MOTU18	1	D05_01	T	100	0	0	0	100
13bp_MOTU17	1	D12_06	T	100	0	0	0	100
13bp_MOTU16	2	D11_18 D06_15	Pl	100	0	0	0	100
13bp_MOTU15	1	D11_27	C	100	0	0	0	100
13bp_MOTU14	1	D06_23	Pris	100	0	0	0	100
13bp_MOTU13	1	D05_02	An	100	0	0	0	100

13bp_MOTU12	1	D05_16	T	100	0	0	0	100
13bp_MOTU11	2	D04_07 D06_08	Pl	100	0	0	0	100
13bp_MOTU10	2	D12_03 D12_04	T	100	0	0	0	100
13bp_MOTU09	1	D04_10	Pl	100	0	0	0	100
13bp_MOTU08	3	D11_19 D11_02 D11_07	Pl	100	0	0	0	100
13bp_MOTU07	1	D12_01	Pl	100	0	0	0	100
13bp_MOTU06	1	D06_17	Pl	100	0	0	0	100
13bp_MOTU05	3	D05_10 D05_15 D04_20	Pl	100	0	0	0	100
13bp_MOTU04	1	D06_30	Do	100	0	0	0	100
13bp_MOTU03	14	D11_10 D11_04 D11_06 D11_28 D11_13 D11_14 D11_20 D11_08 D11_12 D11_21 D11_22 D11_26 D11_25 D11_15	Pl, An	100	0	0	0	100
13bp_MOTU02	1	D04_14	An	100	0	0	0	100
13bp_MOTU01	2	D05_03 D12_08	Pl	100	0	0	0	100
13bp_MOTU	23			2300	0	0	0	2300
14bp_MOTU23	1	D12_06	T	100	0	0	0	100
14bp_MOTU22	1	D11_03	C	100	0	0	0	100
14bp_MOTU21	1	D04_15	Ce	100	0	0	0	100
14bp_MOTU20	2	D04_07 D06_08	Pl	100	0	0	0	100
14bp_MOTU19	2	D12_04 D12_03	T	100	0	0	0	100
14bp_MOTU18	1	D11_09	Ac	100	0	0	0	100
14bp_MOTU17	1	D05_04	Pl	100	0	0	0	100
14bp_MOTU16	1	D05_05	Pl	100	0	0	0	100
14bp_MOTU15	1	D06_17	Pl	100	0	0	0	100
14bp_MOTU14	1	D05_01	T	100	0	0	0	100
14bp_MOTU13	2	D11_18 D06_15	Pl	100	0	0	0	100
14bp_MOTU12	1	D04_14	An	100	0	0	0	100
14bp_MOTU11	1	D12_01	Pl	100	0	0	0	100
14bp_MOTU10	3	D05_10 D05_15 D04_20	Pl	100	0	0	0	100
14bp_MOTU09	1	D05_16	T	100	0	0	0	100
14bp_MOTU08	1	D05_02	An	100	0	0	0	100
14bp_MOTU07	1	D06_30	Do	100	0	0	0	100
14bp_MOTU06	14	D11_13 D11_10 D11_21 D11_25 D11_26 D11_28 D11_22 D11_14 D11_08 D11_12 D11_15 D11_06 D11_20 D11_04	Pl, An	100	0	0	0	100
14bp_MOTU05	2	D05_03 D12_08	Pl	100	0	0	0	100
14bp_MOTU04	1	D11_27	C	100	0	0	0	100
14bp_MOTU03	3	D11_07 D11_02 D11_19	Pl	100	0	0	0	100
14bp_MOTU02	1	D06_23	Pris	100	0	0	0	100
14bp_MOTU01	1	D04_10	Pl	100	0	0	0	100
14bp_MOTU	23			2300	0	0	0	2300
15bp_MOTU23	1	D04_10	Pl	100	0	0	0	100
15bp_MOTU22	1	D12_06	T	100	0	0	0	100
15bp_MOTU21	1	D06_17	Pl	100	0	0	0	100
15bp_MOTU20	1	D06_30	Do	100	0	0	0	100
15bp_MOTU19	1	D11_03	C	100	0	0	0	100
15bp_MOTU18	1	D05_04	Pl	100	0	0	0	100
15bp_MOTU17	2	D12_03 D12_04	T	100	0	0	0	100
15bp_MOTU16	1	D06_23	Pris	100	0	0	0	100
15bp_MOTU15	1	D05_02	An	100	0	0	0	100
15bp_MOTU14	2	D06_08 D04_07	Pl	100	0	0	0	100
15bp_MOTU13	2	D05_03 D12_08	Pl	100	0	0	0	100

15bp_MOTU12	1	D04_14	An	100	0	0	0	100
15bp_MOTU11	1	D12_01	Pl	100	0	0	0	100
15bp_MOTU10	1	D11_09	Ac	100	0	0	0	100
15bp_MOTU09	3	D11_02 D11_07 D11_19	Pl	100	0	0	0	100
15bp_MOTU08	1	D05_01	T	100	0	0	0	100
15bp_MOTU07	1	D11_27	C	100	0	0	0	100
15bp_MOTU06	1	D05_16	T	100	0	0	0	100
15bp_MOTU05	2	D11_18 D06_15	Pl	100	0	0	0	100
15bp_MOTU04	3	D05_10 D05_15 D04_20	Pl	100	0	0	0	100
15bp_MOTU03	1	D04_15	Ce	100	0	0	0	100
15bp_MOTU02	1	D05_05	Pl	100	0	0	0	100
15bp_MOTU01	14	D11_10 D11_13 D11_12 D11_15 D11_26 D11_14 D11_28 D11_04 D11_22 D11_25 D11_21 D11_20 D11_08 D11_06	Pl, An	100	0	0	0	100
15bp_MOTU	23			2300	0	0	0	2300
16bp_MOTU23	1	D12_06	T	100	0	0	0	100
16bp_MOTU22	1	D11_03	C	100	0	0	0	100
16bp_MOTU21	1	D12_01	Pl	100	0	0	0	100
16bp_MOTU20	1	D05_16	T	100	0	0	0	100
16bp_MOTU19	3	D11_19 D11_02 D11_07	Pl	100	0	0	0	100
16bp_MOTU18	1	D04_14	An	100	0	0	0	100
16bp_MOTU17	1	D06_17	Pl	100	0	0	0	100
16bp_MOTU16	1	D05_05	Pl	100	0	0	0	100
16bp_MOTU15	1	D06_23	Pris	100	0	0	0	100
16bp_MOTU14	1	D06_30	Do	100	0	0	0	100
16bp_MOTU13	1	D05_01	T	100	0	0	0	100
16bp_MOTU12	1	D11_09	Ac	100	0	0	0	100
16bp_MOTU11	1	D11_27	C	100	0	0	0	100
16bp_MOTU10	1	D05_02	An	100	0	0	0	100
16bp_MOTU09	3	D04_20 D05_15 D05_10	Pl	100	0	0	0	100
16bp_MOTU08	1	D04_10	Pl	100	0	0	0	100
16bp_MOTU07	1	D05_04	Pl	100	0	0	0	100
16bp_MOTU06	2	D06_08 D04_07	Pl	100	0	0	0	100
16bp_MOTU05	2	D11_18 D06_15	Pl	100	0	0	0	100
16bp_MOTU04	1	D04_15	Ce	100	0	0	0	100
16bp_MOTU03	2	D12_08 D05_03	Pl	100	0	0	0	100
16bp_MOTU02	2	D12_04 D12_03	T	100	0	0	0	100
16bp_MOTU01	14	D11_15 D11_28 D11_13 D11_22 D11_10 D11_06 D11_12 D11_26 D11_21 D11_25 D11_20 D11_08 D11_04 D11_14	Pl, An	100	0	0	0	100
16bp_MOTU	23			2300	0	0	0	2300
17bp_MOTU23	1	D12_01	Pl	100	0	0	0	100
17bp_MOTU22	1	D04_10	Pl	100	0	0	0	100
17bp_MOTU21	1	D05_05	Pl	100	0	0	0	100
17bp_MOTU20	1	D04_14	An	100	0	0	0	100
17bp_MOTU19	1	D06_30	Do	100	0	0	0	100
17bp_MOTU18	1	D05_02	An	100	0	0	0	100
17bp_MOTU17	1	D11_09	Ac	100	0	0	0	100
17bp_MOTU16	1	D05_16	T	100	0	0	0	100
17bp_MOTU15	2	D12_04 D12_03	T	100	0	0	0	100
17bp_MOTU14	1	D06_17	Pl	100	0	0	0	100
17bp_MOTU13	1	D05_01	T	100	0	0	0	100

17bp_MOTU12	1	D11_27	C	100	0	0	0	100
17bp_MOTU11	1	D12_06	T	100	0	0	0	100
17bp_MOTU10	2	D12_08 D05_03	Pl	100	0	0	0	100
17bp_MOTU09	2	D06_08 D04_07	Pl	100	0	0	0	100
17bp_MOTU08	1	D11_03	C	100	0	0	0	100
17bp_MOTU07	1	D06_23	Pris	100	0	0	0	100
17bp_MOTU06	3	D05_15 D04_20 D05_10	Pl	100	0	0	0	100
17bp_MOTU05	1	D05_04	Pl	100	0	0	0	100
17bp_MOTU04	1	D04_15	Ce	100	0	0	0	100
17bp_MOTU03	14	D11_10 D11_21 D11_20 D11_28 D11_06 D11_13 D11_22 D11_15 D11_12 D11_25 D11_04 D11_14 D11_26 D11_08	Pl, An	100	0	0	0	100
17bp_MOTU02	2	D06_15 D11_18	Pl	100	0	0	0	100
17bp_MOTU01	3	D11_07 D11_19 D11_02	Pl	100	0	0	0	100
17bp_MOTU	23			2300	0	0	0	2300
18bp_MOTU23	1	D05_05	Pl	100	0	0	0	100
18bp_MOTU22	2	D04_07 D06_08	Pl	100	0	0	0	100
18bp_MOTU21	1	D06_17	Pl	100	0	0	0	100
18bp_MOTU20	1	D05_01	T	100	0	0	0	100
18bp_MOTU19	1	D04_14	An	100	0	0	0	100
18bp_MOTU18	1	D04_15	Ce	100	0	0	0	100
18bp_MOTU17	1	D05_16	T	100	0	0	0	100
18bp_MOTU16	1	D12_01	Pl	100	0	0	0	100
18bp_MOTU15	2	D06_15 D11_18	Pl	100	0	0	0	100
18bp_MOTU14	1	D11_09	Ac	100	0	0	0	100
18bp_MOTU13	1	D11_03	C	100	0	0	0	100
18bp_MOTU12	1	D06_30	Do	100	0	0	0	100
18bp_MOTU11	1	D06_23	Pris	100	0	0	0	100
18bp_MOTU10	3	D04_20 D05_15 D05_10	Pl	100	0	0	0	100
18bp_MOTU09	2	D12_03 D12_04	T	100	0	0	0	100
18bp_MOTU08	1	D04_10	Pl	100	0	0	0	100
18bp_MOTU07	1	D12_06	T	100	0	0	0	100
18bp_MOTU06	1	D05_02	An	100	0	0	0	100
18bp_MOTU05	1	D05_04	Pl	100	0	0	0	100
18bp_MOTU04	14	D11_20 D11_08 D11_15 D11_21 D11_13 D11_28 D11_14 D11_12 D11_26 D11_25 D11_22 D11_06 D11_10 D11_04	Pl, An	100	0	0	0	100
18bp_MOTU03	3	D11_07 D11_02 D11_19	Pl	100	0	0	0	100
18bp_MOTU02	1	D11_27	C	100	0	0	0	100
18bp_MOTU01	2	D12_08 D05_03	Pl	100	0	0	0	100
18bp_MOTU	23			2300	0	0	0	2300
19bp_MOTU23	1	D11_27	C	100	0	0	0	100
19bp_MOTU22	1	D06_17	Pl	100	0	0	0	100
19bp_MOTU21	1	D12_06	T	100	0	0	0	100
19bp_MOTU20	1	D04_15	Ce	100	0	0	0	100
19bp_MOTU19	1	D11_03	C	100	0	0	0	100
19bp_MOTU18	1	D05_05	Pl	100	0	0	0	100
19bp_MOTU17	1	D06_23	Pris	100	0	0	0	100
19bp_MOTU16	1	D12_01	Pl	100	0	0	0	100
19bp_MOTU15	3	D11_19 D11_07 D11_02	Pl	100	0	0	0	100
19bp_MOTU14	2	D06_15 D11_18	Pl	100	0	0	0	100
19bp_MOTU13	1	D05_02	An	100	0	0	0	100

19bp_MOTU12	1	D06_30	Do	100	0	0	0	100
19bp_MOTU11	1	D11_09	Ac	100	0	0	0	100
19bp_MOTU10	1	D05_16	T	100	0	0	0	100
19bp_MOTU09	2	D06_08 D04_07	Pl	100	0	0	0	100
19bp_MOTU08	2	D05_03 D12_08	Pl	100	0	0	0	100
19bp_MOTU07	2	D12_03 D12_04	T	100	0	0	0	100
19bp_MOTU06	1	D04_10	Pl	100	0	0	0	100
19bp_MOTU05	3	D05_10 D05_15 D04_20	Pl	100	0	0	0	100
19bp_MOTU04	1	D04_14	An	100	0	0	0	100
19bp_MOTU03	1	D05_01	T	100	0	0	0	100
19bp_MOTU02	1	D05_04	Pl	100	0	0	0	100
19bp_MOTU01	14	D11_13 D11_20 D11_08 D11_26 D11_10 D11_21 D11_22 D11_06 D11_28 D11_25 D11_12 D11_04 D11_14 D11_15	Pl, An	100	0	0	0	100
19bp_MOTU	23			2300	0	0	0	2300

LSU	#	Member Sequences		E	S	J	C	Total
8bp_MOTU36	1	D05_06	Ce	100	0	0	0	100
8bp_MOTU35	1	D05_12	Pl	100	0	0	0	100
8bp_MOTU34	1	D05_04	Pl	100	0	0	0	100
8bp_MOTU33	1	D04_17	Do	100	0	0	0	100
8bp_MOTU32	1	D12_16	Ter	100	0	0	0	100
8bp_MOTU31	1	D12_06	T	100	0	0	0	100
8bp_MOTU30	1	D11_18	Pl	100	0	0	0	100
8bp_MOTU29	2	D12_14 D11_16	Do	100	0	0	0	100
8bp_MOTU28	1	D06_02	Do	100	0	0	0	100
8bp_MOTU27	1	D11_03	C	100	0	0	0	100
8bp_MOTU26	3	D06_01 D06_22 D06_32	Ce	100	0	0	0	100
8bp_MOTU25	1	D12_05	T	100	0	0	0	100
8bp_MOTU24	1	D04_18	Des	100	0	0	0	100
8bp_MOTU23	1	D11_07	Pl	100	0	0	0	100
8bp_MOTU22	2	D12_04 D12_03	T	100	0	0	0	100
8bp_MOTU21	3	D11_05 D11_09 D05_05	Ac, Pl	100	0	0	0	100
8bp_MOTU20	2	D06_18 D04_08	Prio	100	0	0	0	100
8bp_MOTU19	1	D12_01	Pl	100	0	0	0	100
8bp_MOTU18	1	D06_03	Ty	100	0	0	0	100
8bp_MOTU17	3	D04_20 D05_07 D05_02	Do, Pl, An	100	0	0	0	100
8bp_MOTU16	3	D04_23 D12_11 D05_03	Pl	100	0	0	0	100
8bp_MOTU15	1	D12_15	Do	100	0	0	0	100
8bp_MOTU14	7	D04_15 D06_29 D05_14 D06_05 D04_19 D12_10 D04_09	Ce, Met	100	0	0	0	100
8bp_MOTU13	2	D06_20 D06_06	Pl	100	0	0	0	100
8bp_MOTU12	4	D05_11 D04_07 D12_08 D06_08	Pl	100	0	0	0	100
8bp_MOTU11	5	D11_29 D11_23 D11_11 D04_12 D04_05	Do	100	0	0	0	100
8bp_MOTU10	7	D11_24 D12_02 D04_11 D04_04 D04_06 D11_01 D05_15	Do, Mes, Pro	100	0	0	0	100
8bp_MOTU09	1	D06_13	Ty	100	0	0	0	100
8bp_MOTU08	6	D12_12 D06_25 D06_19 D04_21 D06_09 D06_27	Do	100	0	0	0	100
8bp_MOTU07	3	D06_21 D04_22 D06_11	Ter	100	0	0	0	100
8bp_MOTU06	5	D04_01 D06_26 D06_14 D06_30 D06_24	Do	100	0	0	0	100
8bp_MOTU05	1	D11_27	C	100	0	0	0	100

8bp_MOTU04	1	D04_10	Pl	100	0	0	0	100
8bp_MOTU03	4	D04_03 D04_16 D04_02 D04_24	Prio	100	0	0	0	100
8bp_MOTU02	16	D11_20 D11_28 D11_13 D11_21 D11_04 D11_12 D11_14 D11_19 D11_15 D06_15 D11_22 D11_08 D11_26 D11_25 D11_06 D11_10	Pl, An	100	0	0	0	100
8bp_MOTU01	5	D11_17 D06_17 D05_10 D04_14 D05_13	Ac, An, Pl	100	0	0	0	100
8bp_MOTU	36			3600	0	0	0	3600
9bp_MOTU37	1	D11_07	Pl	100	0	0	0	100
9bp_MOTU36	1	D12_16	Ter	100	0	0	0	100
9bp_MOTU35	1	D12_01	Pl	100	0	0	0	100
9bp_MOTU34	1	D05_12	Pl	100	0	0	0	100
9bp_MOTU33	1	D11_03	C	100	0	0	0	100
9bp_MOTU32	1	D12_15	Do	100	0	0	0	100
9bp_MOTU31	1	D04_18	Des	100	0	0	0	100
9bp_MOTU30	1	D05_06	Ce	100	0	0	0	100
9bp_MOTU29	2	D12_03 D12_04	T	100	0	0	0	100
9bp_MOTU28	1	D04_17	Do	100	0	0	0	100
9bp_MOTU27	2	D06_06 D06_20	Pl	100	0	0	0	100
9bp_MOTU26	3	D06_32 D06_22 D06_01	Ce	100	0	0	0	100
9bp_MOTU25	2	D06_18 D04_08	Prio	100	0	0	0	100
9bp_MOTU24	1	D06_13	Ty	100	0	0	0	100
9bp_MOTU23	3	D12_08 D04_07 D05_11	Pl	16	0	84	0	100
9bp_MOTU22	1	D12_06	T	100	0	0	0	100
9bp_MOTU21	1	D05_04	Pl	100	0	0	0	100
9bp_MOTU20	1	D11_18	Pl	100	0	0	0	100
9bp_MOTU19	1	D04_10	Pl	100	0	0	0	100
9bp_MOTU18	1	D06_03	Ty	100	0	0	0	100
9bp_MOTU17	3	D04_22 D06_21 D06_11	Ter	100	0	0	0	100
9bp_MOTU16	1	D06_08	Pl	16	0	84	0	100
9bp_MOTU15	2	D12_14 D11_16	Do	100	0	0	0	100
9bp_MOTU14	3	D11_09 D11_05 D05_05	Ac, Pl	100	0	0	0	100
9bp_MOTU13	1	D11_27	C	100	0	0	0	100
9bp_MOTU12	3	D05_03 D12_11 D04_23	Pl	100	0	0	0	100
9bp_MOTU11	3	D05_07 D04_20 D05_02	Do, Pl, An	100	0	0	0	100
9bp_MOTU10	1	D12_05	T	100	0	0	0	100
9bp_MOTU09	5	D04_05 D04_12 D11_23 D11_11 D11_29	Do	100	0	0	0	100
9bp_MOTU08	5	D06_26 D06_30 D06_24 D04_01 D06_14	Do	100	0	0	0	100
9bp_MOTU07	6	D04_21 D06_19 D06_25 D06_27 D06_09 D12_12	Do	100	0	0	0	100
9bp_MOTU06	7	D04_11 D04_06 D11_24 D11_01 D04_04 D05_15 D12_02	Do, Mes, Pro, Pl	100	0	0	0	100
9bp_MOTU05	4	D04_24 D04_16 D04_02 D04_03	Prio	100	0	0	0	100
9bp_MOTU04	1	D06_02	Do	100	0	0	0	100
9bp_MOTU03	5	D05_13 D11_17 D05_10 D04_14 D06_17	Ac, An, Pl	99	0	1	0	100
9bp_MOTU02	16	D11_08 D11_12 D11_06 D11_28 D11_25 D11_26 D11_21 D11_10 D11_20 D11_14 D11_19 D11_13 D11_04 D11_15 D06_15 D11_22	Pl, An	100	0	0	0	100
9bp_MOTU01	7	D12_10 D05_14 D04_09 D06_05 D06_29	Ce,	100	0	0	0	100

		D04_15 D04_19	Met					
9bp_MOTU	37			3531	0	169	0	3700
10bp_MOTU36	1	D06_02	Do	100	0	0	0	100
10bp_MOTU35	1	D04_10	Pl	100	0	0	0	100
10bp_MOTU34	1	D11_03	C	100	0	0	0	100
10bp_MOTU33	1	D12_01	Pl	100	0	0	0	100
10bp_MOTU32	1	D11_27	C	100	0	0	0	100
10bp_MOTU31	1	D12_15	Do	100	0	0	0	100
10bp_MOTU30	1	D05_06	Ce	100	0	0	0	100
10bp_MOTU29	1	D04_18	Des	100	0	0	0	100
10bp_MOTU28	1	D06_03	Ty	100	0	0	0	100
10bp_MOTU27	2	D11_16 D12_14	Do	100	0	0	0	100
10bp_MOTU26	3	D04_23 D12_11 D05_03	Pl	100	0	0	0	100
10bp_MOTU25	3	D06_01 D06_22 D06_32	Ce	100	0	0	0	100
10bp_MOTU24	1	D11_18	Pl	100	0	0	0	100
10bp_MOTU23	1	D12_16	Ter	100	0	0	0	100
10bp_MOTU22	3	D05_05 D11_09 D11_05	Ac, Pl	99	0	1	0	100
10bp_MOTU21	4	D04_16 D04_24 D04_02 D04_03	Prio	100	0	0	0	100
10bp_MOTU20	5	D04_05 D11_11 D11_29 D11_23 D04_12	Do	100	0	0	0	100
10bp_MOTU19	1	D12_06	T	100	0	0	0	100
10bp_MOTU18	1	D05_04	Pl	100	0	0	0	100
10bp_MOTU17	5	D06_17 D05_13 D11_17 D04_14 D05_10	Ac, Pl, An	97	0	3	0	100
10bp_MOTU16	2	D04_08 D06_18	Prio	100	0	0	0	100
10bp_MOTU15	1	D12_05	T	100	0	0	0	100
10bp_MOTU14	4	D05_11 D04_07 D12_08 D06_08	Pl	99	0	1	0	100
10bp_MOTU13	3	D05_07 D05_02 D04_20	Do, Pl, An	99	0	1	0	100
10bp_MOTU12	2	D06_06 D06_20	Pl	98	0	2	0	100
10bp_MOTU11	1	D04_17	Do	100	0	0	0	100
10bp_MOTU10	7	D06_05 D04_15 D12_10 D04_09 D06_29 D04_19 D05_14	Ce, Met	100	0	0	0	100
10bp_MOTU09	2	D12_03 D12_04	T	100	0	0	0	100
10bp_MOTU08	1	D11_07	Pl	100	0	0	0	100
10bp_MOTU07	1	D06_13	Ty	100	0	0	0	100
10bp_MOTU06	6	D04_21 D06_25 D06_09 D06_19 D12_12 D06_27	Do	100	0	0	0	100
10bp_MOTU05	16	D11_15 D11_20 D11_10 D11_06 D11_13 D06_15 D11_14 D11_28 D11_21 D11_12 D11_26 D11_04 D11_25 D11_22 D11_08 D11_19	Pl, An	100	0	0	0	100
10bp_MOTU04	7	D04_11 D04_06 D05_15 D11_01 D12_02 D11_24 D04_04	Do, Mes, Pro, Pl	100	0	0	0	100
10bp_MOTU03	3	D06_11 D04_22 D06_21	Ter	100	0	0	0	100
10bp_MOTU02	1	D05_12	Pl	100	0	0	0	100
10bp_MOTU01	5	D06_24 D04_01 D06_14 D06_26 D06_30	Do	100	0	0	0	100
10bp_MOTU	36			3592	0	8	0	3600

SSU	**#**	**Member Sequences**		**E**	**S**	**J**	**C**	**Total**
13bp_MOTU15	1	D05_14	Met	100	0	0	0	100
13bp_MOTU14	1	D12_06	T	100	0	0	0	100
13bp_MOTU13	1	D12_09	Ty	100	0	0	0	100

13bp_MOTU12	2	D12_03 D12_04	T	100	0	0	0	100
13bp_MOTU11	1	D12_05	T	100	0	0	0	100
13bp_MOTU10	1	D04_18	Des	100	0	0	0	100
13bp_MOTU09	1	D05_16	T	100	0	0	0	100
13bp_MOTU08	3	D05_13 D11_09 D11_05	Ac	100	0	0	0	100
13bp_MOTU07	2	D11_03 D11_27	C	100	0	0	0	100
13bp_MOTU06	1	D05_01	T	100	0	0	0	100
13bp_MOTU05	4	D12_16 D04_22 D06_11 D06_21	Ter	100	0	0	0	100
13bp_MOTU04	10	D06_05 D05_06 D06_22 D04_15 D04_09 D12_10 D04_19 D06_32 D06_29 D06_01	Ce	100	0	0	0	100
13bp_MOTU03	42	D11_20 D11_04 D11_02 D04_07 D06_20 D05_03 D11_19 D06_15 D11_12 D11_21 D12_01 D05_08 D06_17 D11_28 D04_10 D05_12 D06_06 D11_06 D12_08 D06_08 D05_11 D11_07 D11_25 D04_14 D11_13 D11_10 D11_22 D12_11 D04_13 D11_08 D04_20 D05_15 D11_26 D04_23 D05_02 D05_04 D11_14 D11_17 D05_10 D11_15 D11_18 D05_05	Pl, An	100	0	0	0	100
13bp_MOTU02	6	D04_16 D06_18 D04_08 D04_24 D04_03 D04_02	Prio	100	0	0	0	100
13bp_MOTU01	29	D04_17 D12_15 D12_14 D05_09 D04_01 D04_11 D11_16 D12_12 D04_21 D04_05 D06_14 D12_02 D06_27 D06_19 D11_23 D06_24 D06_30 D05_07 D04_04 D11_11 D04_06 D11_29 D06_02 D06_09 D11_01 D06_26 D04_12 D06_25 D11_24	Do, Pro, Mes	96	4	0	0	100
13bp_MOTU	15			1496	4	0	0	1500
14bp_MOTU15	1	D05_01	T	100	0	0	0	100
14bp_MOTU14	1	D12_05	T	100	0	0	0	100
14bp_MOTU13	1	D12_06	T	100	0	0	0	100
14bp_MOTU12	1	D12_09	Ty	100	0	0	0	100
14bp_MOTU11	1	D05_16	T	100	0	0	0	100
14bp_MOTU10	1	D04_18	Des	100	0	0	0	100
14bp_MOTU09	4	D06_11 D12_16 D06_21 D04_22	Ter	100	0	0	0	100
14bp_MOTU08	3	D11_05 D11_09 D05_13	Ac	100	0	0	0	100
14bp_MOTU07	2	D11_03 D11_27	C	100	0	0	0	100
14bp_MOTU06	2	D12_04 D12_03	T	100	0	0	0	100
14bp_MOTU05	10	D06_32 D06_05 D06_22 D06_29 D06_01 D04_09 D05_06 D04_15 D12_10 D04_19	Ce	100	0	0	0	100
14bp_MOTU04	42	D05_10 D11_18 D04_20 D04_13 D06_08 D05_11 D11_08 D12_08 D11_14 D11_20 D04_14 D11_06 D11_12 D06_06 D11_13 D11_19 D05_15 D11_26 D05_04 D11_04 D06_17 D11_17 D11_28 D05_08 D05_03 D11_07 D11_15 D05_05 D05_02 D12_01 D11_22 D06_20 D12_11 D04_07 D04_23 D11_10 D11_02 D06_15 D04_10 D11_25 D11_21 D05_12	Pl, An	100	0	0	0	100
14bp_MOTU03	6	D06_18 D04_08 D04_02 D04_24 D04_16 D04_03	Prio	100	0	0	0	100
14bp_MOTU02	1	D05_14	Met	100	0	0	0	100
14bp_MOTU01	29	D04_17 D12_14 D12_02 D04_01 D04_04 D06_27 D06_30 D11_29 D06_09 D06_19 D05_09 D04_21 D04_12 D11_24 D11_01 D06_26 D11_16 D12_12 D05_07 D06_14 D06_24 D11_23 D04_05 D12_15 D04_06	Do, Pro, Mes	91	9	0	0	100

		D06_25 D11_11 D04_11 D06_02						
14bp_MOTU	15			1491	9	0	0	1500
15bp_MOTU15	1	D12_09	Ty	100	0	0	0	100
15bp_MOTU14	1	D04_18	Des	100	0	0	0	100
15bp_MOTU13	1	D05_14	Met	100	0	0	0	100
15bp_MOTU12	1	D05_01	T	100	0	0	0	100
15bp_MOTU11	2	D12_04 D12_03	T	100	0	0	0	100
15bp_MOTU10	1	D05_16	T	100	0	0	0	100
15bp_MOTU09	3	D11_05 D11_09 D05_13	Ac	100	0	0	0	100
15bp_MOTU08	2	D11_03 D11_27	C	100	0	0	0	100
15bp_MOTU07	6	D04_02 D04_08 D06_18 D04_03 D04_16 D04_24	Prio	100	0	0	0	100
15bp_MOTU06	1	D12_06	T	100	0	0	0	100
15bp_MOTU05	1	D12_05	T	100	0	0	0	100
15bp_MOTU04	29	D06_14 D06_30 D11_29 D06_02 D06_25 D12_12 D04_01 D11_24 D12_15 D12_02 D06_19 D06_26 D04_06 D11_11 D11_16 D04_04 D04_12 D11_23 D05_07 D04_17 D05_09 D04_05 D06_09 D06_27 D06_24 D11_01 D12_14 D04_21 D04_11	Do, Pro, Mes	91	9	0	0	100
15bp_MOTU03	4	D12_16 D06_21 D06_11 D04_22	Ter	100	0	0	0	100
15bp_MOTU02	10	D06_01 D04_19 D06_29 D12_10 D04_15 D04_09 D06_32 D06_22 D05_06 D06_05	Ce	100	0	0	0	100
15bp_MOTU01	42	D06_15 D11_04 D11_25 D05_04 D11_14 D11_07 D06_17 D05_08 D05_11 D12_01 D11_22 D11_13 D12_08 D06_08 D05_12 D05_02 D11_21 D11_15 D11_08 D12_11 D11_17 D11_28 D11_26 D04_07 D06_06 D11_19 D06_20 D05_05 D04_10 D11_18 D05_10 D11_06 D11_10 D04_14 D11_20 D05_03 D04_13 D04_20 D04_23 D11_02 D05_15 D11_12	Pl, An	100	0	0	0	100
15bp_MOTU	15			1491	9	0	0	1500
16bp_MOTU14	1	D12_09	Ty	100	0	0	0	100
16bp_MOTU13	1	D04_18	Des	100	0	0	0	100
16bp_MOTU12	1	D05_14	Met	100	0	0	0	100
16bp_MOTU11	2	D11_27 D11_03	C	100	0	0	0	100
16bp_MOTU10	6	D04_08 D06_18 D04_03 D04_16 D04_02 D04_24	Prio	100	0	0	0	100
16bp_MOTU09	1	D12_05	T	100	0	0	0	100
16bp_MOTU08	2	D12_04 D12_03	T	100	0	0	0	100
16bp_MOTU07	4	D06_21 D12_16 D04_22 D06_11	Ter	100	0	0	0	100
16bp_MOTU06	3	D11_09 D05_13 D11_05	Ac	100	0	0	0	100
16bp_MOTU05	2	D05_16 D05_01	T	100	0	0	0	100
16bp_MOTU04	29	D06_14 D04_01 D05_09 D06_09 D04_05 D06_24 D06_27 D06_02 D11_11 D04_12 D04_04 D04_17 D12_15 D11_23 D11_16 D06_25 D06_26 D12_12 D04_06 D06_19 D04_11 D04_21 D06_30 D05_07 D12_02 D11_29 D12_14 D11_24 D11_01	Do, Pro, Mes	96	4	0	0	100
16bp_MOTU03	1	D12_06	T	100	0	0	0	100
16bp_MOTU02	10	D05_06 D04_19 D06_29 D06_05 D12_10 D04_09 D06_01 D06_32 D04_15 D06_22	Ce	100	0	0	0	100
16bp_MOTU01	42	D11_22 D06_08 D06_06 D05_10 D06_20 D05_15 D11_15 D06_17 D04_14 D05_02	Pl, An	100	0	0	0	100

		D11_06 D04_13 D04_23 D11_17 D11_13 D11_21 D05_05 D12_08 D11_14 D11_02 D05_03 D04_20 D05_08 D11_12 D04_10 D11_07 D11_28 D11_20 D11_08 D05_12 D05_04 D11_04 D11_10 D12_01 D11_26 D04_07 D11_18 D11_19 D12_11 D06_15 D05_11 D11_25						
16bp_MOTU	14			1396	4	0	0	1400
17bp_MOTU14	1	D12_06		100	0	0	0	100
17bp_MOTU13	3	D11_09 D11_05 D05_13		100	0	0	0	100
17bp_MOTU12	2	D12_03 D12_04		100	0	0	0	100
17bp_MOTU11	1	D05_14		100	0	0	0	100
17bp_MOTU10	1	D12_05		100	0	0	0	100
17bp_MOTU09	2	D05_16 D05_01		100	0	0	0	100
17bp_MOTU08	1	D12_09		100	0	0	0	100
17bp_MOTU07	4	D06_21 D06_11 D04_22 D12_16		100	0	0	0	100
17bp_MOTU06	1	D04_18		100	0	0	0	100
17bp_MOTU05	2	D11_03 D11_27		100	0	0	0	100
17bp_MOTU04	10	D06_05 D04_09 D06_22 D04_19 D04_15 D06_01 D05_06 D12_10 D06_32 D06_29		100	0	0	0	100
17bp_MOTU03	29	D12_12 D06_30 D06_24 D06_02 D04_11 D04_21 D04_04 D04_01 D06_27 D11_11 D11_29 D12_02 D05_07 D06_25 D04_06 D06_09 D12_14 D11_01 D06_26 D11_16 D04_05 D05_09 D11_23 D12_15 D04_12 D11_24 D06_14 D04_17 D06_19		95	5	0	0	100
17bp_MOTU02	6	D06_18 D04_16 D04_02 D04_24 D04_03 D04_08		100	0	0	0	100
17bp_MOTU01	42	D11_15 D04_23 D11_02 D06_17 D05_15 D11_28 D11_07 D05_08 D11_26 D06_08 D04_13 D05_12 D12_01 D06_20 D12_11 D05_04 D11_18 D11_14 D11_08 D11_25 D11_21 D11_13 D11_10 D05_05 D11_19 D12_08 D11_20 D05_10 D04_14 D11_22 D11_06 D04_20 D05_03 D05_11 D06_15 D04_07 D11_12 D04_10 D05_02 D11_04 D11_17 D06_06		100	0	0	0	100
17bp_MOTU	14			1395	5	0	0	1400
18bp_MOTU14	1	D04_18	Des	100	0	0	0	100
18bp_MOTU13	1	D12_05	T	100	0	0	0	100
18bp_MOTU12	1	D05_14	Met	100	0	0	0	100
18bp_MOTU11	1	D12_09	Ty	100	0	0	0	100
18bp_MOTU10	3	D05_13 D11_09 D11_05	Ac	100	0	0	0	100
18bp_MOTU09	2	D05_01 D05_16	T	100	0	0	0	100
18bp_MOTU08	1	D12_06	T	100	0	0	0	100
18bp_MOTU07	10	D06_22 D04_19 D04_09 D06_32 D06_05 D06_29 D06_01 D12_10 D05_06 D04_15	Ce	100	0	0	0	100
18bp_MOTU06	4	D12_16 D06_21 D04_22 D06_11	Ter	100	0	0	0	100
18bp_MOTU05	2	D11_27 D11_03	C	100	0	0	0	100
18bp_MOTU04	6	D04_08 D04_24 D04_16 D06_18 D04_03 D04_02	Prio	100	0	0	0	100
18bp_MOTU03	2	D12_04 D12_03	T	100	0	0	0	100
18bp_MOTU02	29	D12_12 D04_12 D06_02 D11_16 D06_19 D11_11 D05_07 D06_26 D04_21 D06_30 D12_15 D04_05 D06_14 D06_09 D06_24 D11_01 D06_25 D04_06 D04_17 D11_29	Do, Pro, Mes	100	0	0	0	100

		D04_04 D06_27 D04_01 D11_23 D05_09 D04_11 D12_02 D11_24 D12_14						
18bp_MOTU01	42	D11_15 D11_28 D12_01 D05_03 D05_05 D11_26 D06_08 D11_17 D04_14 D05_08 D11_25 D04_23 D11_06 D11_04 D12_11 D12_08 D11_08 D06_15 D06_17 D04_20 D06_06 D04_13 D11_10 D06_20 D11_19 D11_14 D04_07 D11_02 D11_12 D11_21 D05_12 D04_10 D05_15 D11_20 D05_10 D05_11 D05_02 D05_04 D11_18 D11_22 D11_13 D11_07	Pl, An	100	0	0	0	100
18bp_MOTU	14			1400	0	0	0	1400

Nematology, 2006, Vol. 8(3), 367-376

DESS: a versatile solution for preserving morphology and extractable DNA of nematodes

Melissa YODER [1,*], Irma TANDINGAN DE LEY [1], Ian Wm KING [1], Manuel MUNDO-OCAMPO [1], Jenna MANN [2], Mark BLAXTER [2], Larisa POIRAS [3] and Paul DE LEY [1]

[1] *Department of Nematology, University of California, Riverside, CA 92521, USA*
[2] *School of Biological Sciences, University of Edinburgh, Institute of Evolutionary Biology Ashworth Laboratories, King's Buildings, Edinburgh EH9 3JT, UK*
[3] *Institute of Zoology, Academy of Sciences, Str. Academie 1, 2028 Chisinau, Moldova*

Received: 7 November 2005; revised: 13 March 2006
Accepted for publication: 15 March 2006

Summary – A solution containing dimethyl sulphoxide, disodium EDTA, and saturated NaCl (abbreviated here as DESS) was tested for various applications in the preservation of nematodes for combined morphological and molecular analyses. The solution can be used to preserve individual nematodes, nematode extracts, or entire soil/sediment samples. Preserved material can be easily stored for months at room temperature, shipped by mail, or carried in luggage. Morphological features are usually well preserved; specimen quality being comparable to formalin-based fixatives and much better than ethanol fixation. Specimens can be transferred to glycerin with little or no modification of traditional protocols. Unlike formalin-preserved material, routine PCR can be performed on individual specimens after any of these procedures with success rates and amplification sizes comparable to PCR of fresh specimens. At this point we have no data on long-term preservation quality. Nevertheless, DESS solution clearly enhances and simplifies a wide range of nematological studies due to its combined suitability for morphological and molecular analyses, as well as its less hazardous chemical properties.

Keywords – dimethyl sulphoxide, EDTA, molecular, PCR, protocol, SEM.

With the discovery of the polymerase chain reaction (PCR) and the continual reduction of costs and time to amplify DNA, molecular studies have become widespread in all fields of biological research. Although DNA is a very informative molecule, it usually requires rapid intervention to avoid its degradation into small fragments by active nucleases. Specifically, at least one of three factors must be adjusted rapidly to inactivate these nucleases: temperature, pH, or salt concentration (Dixon & Webb, 1979). Dessauer *et al.* (1996) found cryopreservation to be the most effective method for long-term preservation of DNA. This is a useful method when working in a laboratory, but equipment needed to reach temperatures as low as −70°C is difficult to transport to and from the field. Samples must therefore be treated on-site with a preservative that will stop nuclease activity. In nematology, ethanol and formalin have been most frequently used or tested to preserve nematode DNA (Thomas *et al*., 1997; Schander & Halanych, 2003; Roubtsova *et al.*, 2005). However, neither of these solutions combines the properties of an ideal fixative, *i.e*., adequate preservation of both DNA and morphology, straightforward handling in terms of transportation to and from the field, easy storage after collection of specimens, plus minimal hazard in terms of flammability or toxicity. As a result, a sample must often be split into subsamples for preservation of DNA in ethanol and morphology in formalin. This approach is less than ideal because it allows for discrepancies between subsamples in species composition and also because it precludes obtaining combined sequence and morphology data from single nematodes.

Several studies have compared and described the effects of different preservatives on DNA (Greer *et al.*, 1991; Seutin *et al.*, 1991; Dillon *et al*., 1996; Holzmann & Pawlowski, 1996; Miller & Hook, 1996; Thomas *et al.*, 1997; Toe *et al*., 1997; Dawson *et al.*, 1998; Kilpatrick, 2002). Seutin *et al*. (1991) suggested the use of a DMSO/EDTA/saturated NaCl solution for preserving DNA from avian tissues at room temperature. Dawson *et al*. (1998) subsequently showed that this solution allows

* Corresponding author, e-mail: melissay@ucr.edu

© *Koninklijke Brill NV, Leiden, 2006*
Also available online - www.brill.nl

preservation of both physical structures and high molecular weight DNA for up to 6 months without freezing. Kilpatrick (2002) tested long-term DNA preservation and determined that high molecular weight DNA was amplifiable for up to 2 years in this solution. We report here on the various uses of this solution (henceforth abbreviated as DESS solution) for combined preservation of nematode DNA and morphology.

Materials and methods

Preservation of extracts of terrestrial and marine nematodes

Soil samples were collected from the UCR Botanic Gardens and nematodes were extracted by modified Baermann technique (Schindler, 1961). In addition, a freshly extracted sample from Kern County, CA, USA (supplied by Dr Michael McKenry) was also used for initial testing. Individuals were picked for video capture and editing (VCE) (De Ley & Bert, 2002), PCR, and sequencing. Each specimen was mounted temporarily under a cover glass on ringed fluorescence slides in a drop of deionised water. The temporary mount was examined on an Olympus® BX51 microscope with differential interference contrast optics and the most important body parts were imaged *via* a Matrox RTMac on a Macintosh G4 PowerPC or Sony™ Handycam® HDR-HC1K Digital HD Video Camera Recorder using Apple iMovie HD version 5.0.2, as multifocal images. Multifocal vouchers, sample information, and DNA sequences (when available) were deposited in NemATOL (http://nematol.unh.edu/) for all specimens analysed.

The remaining nematodes were preserved in bulk by replacing water with DESS solution containing 20% dimethyl sulphoxide (DMSO) and 0.25 M disodium EDTA, saturated with NaCl, pH 8.0 (Seutin *et al*., 1991) (http://nematol.unh.edu/Method-Protocol/DMSO_protocol.htm). This was done by pouring the extract over a 500 mesh sieve (25 μm opening) allowing most of the water to drain, then replacing with DESS solution as the contents of the sieve were collected into a vial. Preserved samples were stored at room temperature for 2 weeks, 7 weeks, and 7 months at which time individual nematodes were picked from the DESS solution, washed in deionised water, and video captured prior to PCR and sequencing.

Marine nematodes from Solana Beach, CA, USA, were obtained from the intertidal zone or from kelp holdfasts. Sediment was placed in a bucket with three times its volume of 3% artificial seawater, stirred by hand, decanted over a 500 (25 μm) mesh sieve several times, and preserved in DESS solution as described above. Holdfasts were washed several times in a tray with 3% artificial seawater prior to decanting and preservation. Three days later, individual nematodes were picked from the DESS solution and video captured prior to PCR and sequencing.

Safe transportation of samples by air or by mail

Soil, lichen, and moss samples were collected from various sites in Moldova during an exchange visit organised by the National Academy of Sciences International Research Experiences for Undergraduates (INTREU) programme between the University of California campuses at Riverside and Davis, Brigham Young University, and Moldova State University. Soil samples were processed using a series of 60 (250 μm), 200 (75 μm), and 500 (25 μm) mesh sieves and extracts placed on a modified Baermann for 3 days. Mosses and lichens were placed on a modified Baermann for 5 days to extract all nematodes. Each sample was split into subsamples for preservation in 5% formalin and DESS solution. All subsamples were stored at room temperature from a few days up to 2 weeks and then brought to UCR by air in checked luggage. After 24 h of travelling, samples were stored at room temperature for between 2 days and 6 months. Subsamples with DESS solution were then examined under the dissection microscope and nematodes were picked for VCE, PCR, and sequencing.

Several soil sample extracts preserved in DESS solution were also exchanged by mail between Dr Mark Blaxter's laboratory in Edinburgh, UK, and our laboratories at UCR. Samples took 3-4 weeks to arrive when nematodes were picked for VCE, PCR, and sequencing. Samples sent to the Blaxter laboratory were left at room temperature for up to 9 weeks, after which individuals were rinsed in sterile tap water before PCR and sequencing.

Preparation of glycerin mounts

To test if DESS solution can preserve DNA and replace formalin in the preparation of glycerin mounts with the Seinhorst (1959) method as modified by De Grisse (1969), preserved nematodes from the UCR Botanic Gardens and Solana Beach, California, were first rinsed with purified water to remove any debris. A glass cavity block containing the nematode extract in purified water was then placed in an airtight jar containing 1.25 cm deep volume

of 96% ethanol and left overnight in an incubator set to 40°C. The glass block was removed from the jar the next morning, filled to the brim with five parts glycerol and 95 parts 96% ethanol solution, and left at 40°C with two-thirds of its cavity covered by a glass square. Gradual transition to glycerin was achieved by adding more of the glycerol:ethanol (5:95) solution every few hours. The next day, individual nematodes were mounted on glass slides which were set aside for between one week and five months. At the end of the allotted time period, one slide was broken open and the nematodes were washed three times in purified deionised water to remove excess glycerin. Individuals were then subjected to VCE, PCR, and sequencing.

Bulk preservation of soil samples before nematode extraction

To investigate if the DESS solution could also replace formalin as a bulk preservative, DESS solution was used to preserve entire samples of soil or sediment obtained from the UCR campus and the Salton Sea, CA, USA. DESS solution was added at a ratio of 3:1 to each sample and mixed thoroughly. Samples were stored at room temperature for 1-2 weeks. Then 200 g of substrate and DESS solution were weighed from each sample, mixed with water, and decanted through a series of 80 (180 μm) and 500 (25 μm) mesh sieves for a minimum of three times per sample. The substrate remaining after decantation was centrifuged at 1450 *g* for 5 min and the supernatant recovered and set aside. Ludox® TM-50 Colloidal Silica (Grace Davison W.R. Grace & Co.-Conn., Columbia, MA, USA) at a 2:3 dilution in purified water was added and mixed thoroughly to each substrate. Samples were centrifuged again at 1450 *g* for 15 min and the supernatant recovered and set aside. Ludox® was added again, the sample was mixed and centrifuged as above a minimum of three times per sample. The supernatant set aside from each centrifugation was then rinsed with deionised water in a 500 (25 μm) mesh sieve and placed back into DESS solution for examination. Nematodes were picked from both extract fractions and recorded by VCE before PCR and sequencing.

Preservation of an arthropod host before nematode extraction

A 6 cm long polydesmid millipede (species and genus as yet unidentified) was captured in the Amazon forest of Peru, killed, stored in 30 ml of DESS solution, and shipped at ambient temperature to Edinburgh, UK. It was stored at −20°C in DESS for 1 month before dissection. The specimen was thawed, head and tail were removed, and the body was sliced through every second diplosegment. Gut contents were washed into sterile tap water and examined for nematode parasites under a dissection microscope. A subset of the recovered nematodes, selected for morphological disparity, were picked to microscope slides and imaged on a Zeiss Axiovert 35 microscope with Open*lab*™ (Impro*vision*®, Lexington, MA, USA) digital imaging system. Selections of images are available on the Internet for remote diagnosis (http://nemhelix.cap.ed.ac.uk/mpl/Peru_Nematodes/peruvian_millipede_nematod.html). All nematodes recovered from the millipede gut were used for PCR and sequencing.

PCR and sequencing of target loci

Each recorded specimen was recovered intact from the temporary mount, cut into two pieces in 20 μl of Worm Lysis Buffer (50 mM KCl, 10 mM Tris-Cl pH 8.3, 2.5 mM $MgCl_2$, 0.45% NP40, and 0.45% Tween 20, as described in Williams *et al.*, 1992), transferred to a microcentrifuge tube containing 2 μl of Proteinase K (60 μg ml^{-1}) for digestion, and stored at −80°C. PCR amplification was subsequently performed on either the D2D3 domain of the large subunit (LSU) or 18S (small subunit or SSU) rDNA gene. A 25 μl reaction contained 2.5 μl of genomic DNA as template, 2.5 μl of 10× reaction buffer with $MgCl_2$, dNTP-mix at 0.2 mM each, 0.4 μM each of primer (A (5′-AAA GAT TAA GCC ATG CAT G-3′) and G18S4 or 18P (Blaxter *et al*., 1998; Tandingan De Ley *et al*., 2002) for SSU; D2Ab and D3b (De Ley *et al*., 1999) for D2D3), and 1 unit of DyNAzyme EXT DNA polymerase (New England Biolabs®, Ipswich, MA, 01938, USA). PCR conditions were: denaturation at 94°C for 30 s, annealing at 55°C for 1 min, and extension at 72°C for 2 min, followed by polymerisation for 7 min at 72°C for 35 cycles. PCR products were separated on a 1% agarose gel stained with 0.0003% ethidium bromide with 1 kbp DNA ladder (Promega, Madison, WI, USA) as size markers. Positive products were cleaned with QIAquick® PCR Purification Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. Sequencing was performed using a 96-capillary ABI 3730xl at the UCR Core Instrumentation Facility. Sequences were assembled using GeneTool 2.0 (Biotools, Edmonton, AB, Canada) and compared with published sequences in GenBank by means of BLAST search (Altschul *et al*., 1997).

Specimens used for PCR amplification in Edinburgh were picked and digested in 20 μl (40 μl for large nematodes) of 0.25 M NaOH. Digests were incubated overnight at room temperature, heated to 95°C for 3 min, neutralised (4 μl 1 M HCl, 10 μl 0.5 M Tris-HCl (pH 8.0), 5 μl 2% Triton X-100 per digest) and then heated again at 95°C for 3 min (Floyd *et al*., 2002). PCR amplification was performed on the SSU rDNA gene using 2-4 μl of the neutralised lysate, primer SSU_F_04 (sequence identical with G18S4) and SSU_R_26 (5′-CAT TCT TGG CAA ATG CTT TCG-3′). PCR conditions were: denaturation at 94°C for 5 min followed by 35 cycles of 94°C for 1 min, annealing at 55°C for 90 s and extension at 72°C for 2 min, followed by a final extension stage at 72°C for 10 min. PCR products were visualised by separation on a 1.5% agarose gel stained with 0.0002% ethidium bromide with 1 Kbp DNA ladder (Invitrogen™, Carlsbad, CA, USA) as size markers. Positive products were cleaned using exonuclease I (New England Biolabs®, Ipswich, MA, USA) and shrimp alkaline phosphatase (USB®, Cleveland, OH, USA) following the Wellcome Trust Sanger Institute's (http://www.sanger.ac.uk/) protocol and sequenced using the reverse primer SSU_R_09 (5′-AGC TGG AAT TAC CGC GGC TG-3′). Sequencing was performed in a 48-capillary ABI 3730, and assembled sequences were compared with published sequences in GenBank by means of BLAST search.

Preparation of nematodes for scanning electron microscopy

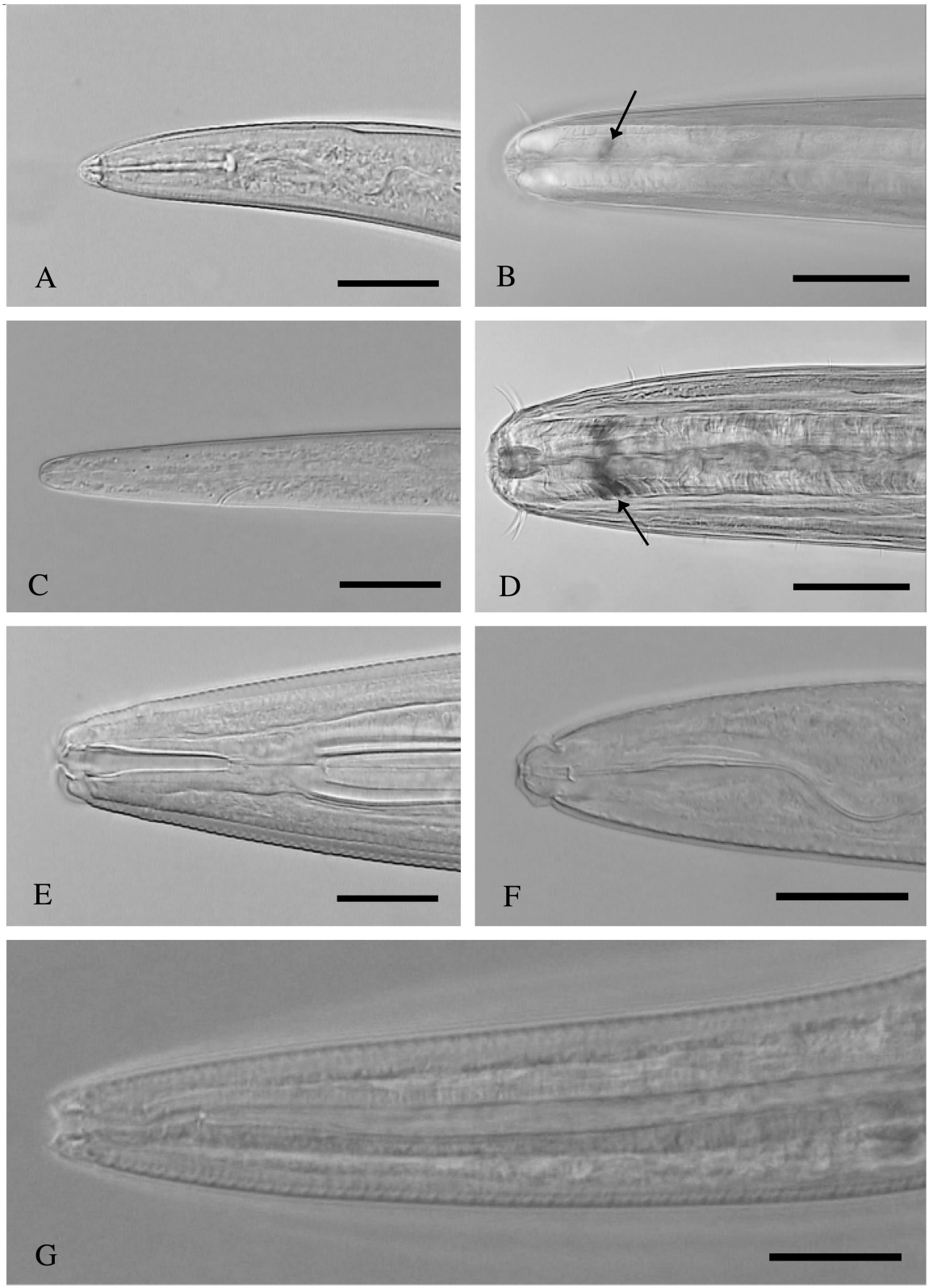
Marine sediment samples were collected from the intertidal zone at Redondo Beach, CA, USA and Mazatlan, Sinaloa, Mexico. Nematodes were extracted from the marine sediments as described above, divided into two subsamples, one preserved in DESS solution while the other was fixed in 5% formalin. Preserved nematodes from Redondo Beach were left at room temperature for 24 h while those from the Mazatlan samples were left for 6 weeks before SEM preparation. Nematodes were rinsed with several changes of deionised water followed by 0.1 M phosphate buffer (pH 7.0), hand picked, and transferred to a BEEM® capsule (Ted Pella, Redding, CA, USA). After post-fixation in 4% OsO_4, specimens were rinsed with three changes of cold (4°C) 0.1 M phosphate buffer (pH 7.2) within a 15 min period, dehydrated through a graded series of 100% absolute ethanol, and critical point dried using an Tousimis Autosamdri® -815 critical point drier. Dried nematodes were mounted on the surface of double-sided copper tape attached to aluminium stubs and sputter coated with a 25 nm layer of gold palladium in a Cressington 108 Auto. Specimens were observed using a Philips® XL30-FEG Scanning Electron Microscope (SEM) operated at 10 kV.

Results and discussion

All individuals video captured prior to PCR were identified to genus using the recorded multifocal images. Nematodes from 26 families consisting of 38 genera were isolated for VCE and PCR. Internal and external structures were usually well preserved in specimens left in DESS solution for periods between 3 days to 7 months (Fig. 1). Eyespot and muscle pigmentation in pharyngeal tissues of marine nematodes from kelp holdfasts were still visible after preservation in DESS solution (Fig. 1B, D, arrows). Samples that were bulk preserved in DESS solution and transported internationally also showed well-preserved morphology. Morphological preservation quality in glycerin mounts after 1 week was comparable overall to formalin-fixed specimens and, in the case of infective juveniles of *Steinernema*, it was noticeably better, presumably due to the permeabilising properties of DMSO (Fig. 1C). After 5 months in glycerin mounts, specimens showed no signs of morphological deterioration and structures remained well preserved (Fig. 1G). Hence, DESS solution is a good short-term preservative for morphology.

Direct observation of specimens immediately after the addition of DESS solution revealed that nematodes became distorted due to collapse of the body, especially for species from freshwater environments. However, in the course of the next few minutes to hours, specimens

Fig. 1. *Nematodes preserved in DESS solution used for VCE, PCR and sequencing. A: 7M5G5* – Helicotylenchus *preserved in DESS solution for 7 weeks; B: 1I17H5* – Enoplus *with eyespot pigmentation still visible (arrow); C: 8M4H5* – Steinernema *juvenile preserved in DESS solution, transferred to glycerin, and mounted on a glass slide for 1 week; D: 2I15H5* – Enoplus *preserved in DESS solution, transferred to glycerin and mounted on a glass slide, showing pigmentation in pharyngeal tissues (arrow); E: 2M25G5* – Plectus *transported back to the USA from Moldova* via *air; F: 13M17H5* – Leptonchus *extracted by Ludox® TM-50 centrifugation; G: 19M9A6* – Eucephalobus *preserved in DESS solution, transferred to glycerin, and mounted on a glass slide for 5 months, captured on a Sony™ Handycam Digital HD Video Camera Recorder. (Scale bars: A, C, E-G = 10 μm; B, D = 30 μm.)*


A
B
C
D
E
F
G

slowly reinflated and eventually resumed their original turgor and appearance. Presumably this phenomenon is caused by the severe initial osmotic shock, followed by a slower return to osmotic balance under the permeabilising effects of the DMSO in the solution. Dimethyl sulphoxide is used to permeate the tissues and membranes while transporting non-ionised molecules of low molecular weight into tissues, membranes, and, eventually, the cells of an organism (Jacobs, 1971). Therefore, the disodium EDTA and sodium chloride in the DESS solution are transported quickly into the cells of the specimen with the help of DMSO, inactivating enzymes that degrade DNA.

DNA amplification and sequencing of individuals was successful with no need to change existing PCR protocols. Amplification of DNA fragments from 800 bp to 1800 bp was 80% successful and sequencing success from these amplicons was greater than 90% for all nematodes preserved in DESS solution for 3 days to 7 months. DNA amplification from D2D3 and 18S produced bright to moderate bands for all terrestrial and marine nematodes preserved in DESS solution. Samples transported internationally, samples bulk preserved in DESS solution, and samples transferred to glycerin for permanent mounts showed similar results for DNA amplification (Fig. 2) and sequencing as those above. Individuals used for PCR and sequencing from permanent mounts 5 months after mounting also showed bright to moderate bands for both D2D3 and 18S products. Average sequence read lengths for D2D3 and 18S PCR products were 750 bp and 1730 bp, respectively, which is comparable to sequences obtained from freshly lysed nematodes. Preserved samples used for DNA amplification and sequencing in Edinburgh yielded similar results with 73% and 97% success, respectively. Amplification produced bright to moderate bands *ca* 900 bp long and sequencing reads extended to the end of the amplified product (*ca* 500 bp). These results differ in several respects from those obtained from samples preserved in ethanol or formalin. Holzmann and Pawlowski (1996) concluded that 70% ethanol preserves DNA fragments up to 1300 bp, but amplification produced only weak bands. In our own experience, ethanol-preserved material performs quite inconsistently, perhaps due to problems with impurities or vapour substitution with air humidity. DNA amplification from formalin-preserved material is no better, resulting in DNA amplification of fragments 400 bp or less (Thomas *et al*., 1997; Dorris *et al.*, 2002; Schander & Halanych, 2003; Roubtsova *et al.*, 2005). Formalin-preserved DNA is also prone to nucleotide substitutions and problems with primer annealing caused by denaturing double-stranded DNA due to the use of hot formalin (Schander & Halanych, 2003). DMSO is known to perturb protein structure (Rammler, 1971) and can inhibit *Taq* polymerase activity during PCR (Gelfland, 1989). However, there is no evidence to suggest that nucleotide substitutions occur due to preservation in DESS solution. It is recommended that specimens preserved in DESS solution be washed thoroughly with purified deionised water before being used for DNA amplification and sequencing to avoid excess salts and residual DMSO in a PCR reaction. Consequently, the DESS solution is a better DNA preservative because it combines consistent PCR performance with the ability to amplify and sequence high molecular weight DNA.

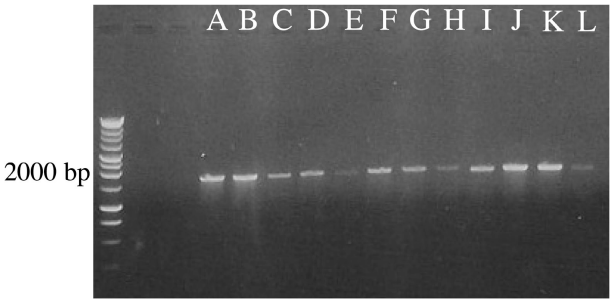


Fig. 2. *SSU PCR products for nematodes preserved in DESS solution. A-E, G-J: Nematodes preserved in DESS solution, transferred to glycerin, and mounted on glass slides for 1 week. A: 1M4H5* – Steinernema *juvenile; B: 2M4H5* – Steinernema *juvenile; C: 3M4H5* – Steinernema *juvenile; D: 4M4H5* – Steinernema *juvenile; E: 5M4H5* – Aphelenchoides *female; G: 6M4H5* – Steinernema *juvenile; H: 7M4H5* – Steinernema *juvenile; I: 8M4H5* – Steinernema *juvenile; J: 9M4H5* – Aphelenchoides *female. F, K-L: Preserved nematodes shipped from the Blaxter laboratory in Edinburgh, UK. F: 13M4H5 – Tylenchidae male; K: 11M4H5* – Plectus *female; L: 12M4H5* – Plectus *female.*

DESS can also be used to preserve parasitic nematodes *in situ* inside hosts. From a single polydesmid millipede, 28 individual gut-parasitic nematodes were isolated, ranging from *ca* 0.5 mm to 4 mm in length. The specimens retained internal and external morphological characters (Fig. 3), and Dr David Hunt of CABI examined the digital images obtained. Notably, even fine details of decoration on the impressive spines (Fig. 3E) and other cuticular decoration of the specimens were visible. Five putative morphospecies were identified, belonging to the Hethidae, Rhigonematidae and Carnoyidae, all part of the Rhigonematomorpha (*sensu* De Ley & Blaxter, 2002). Rhigonematomorphs are a clade of nematodes related to

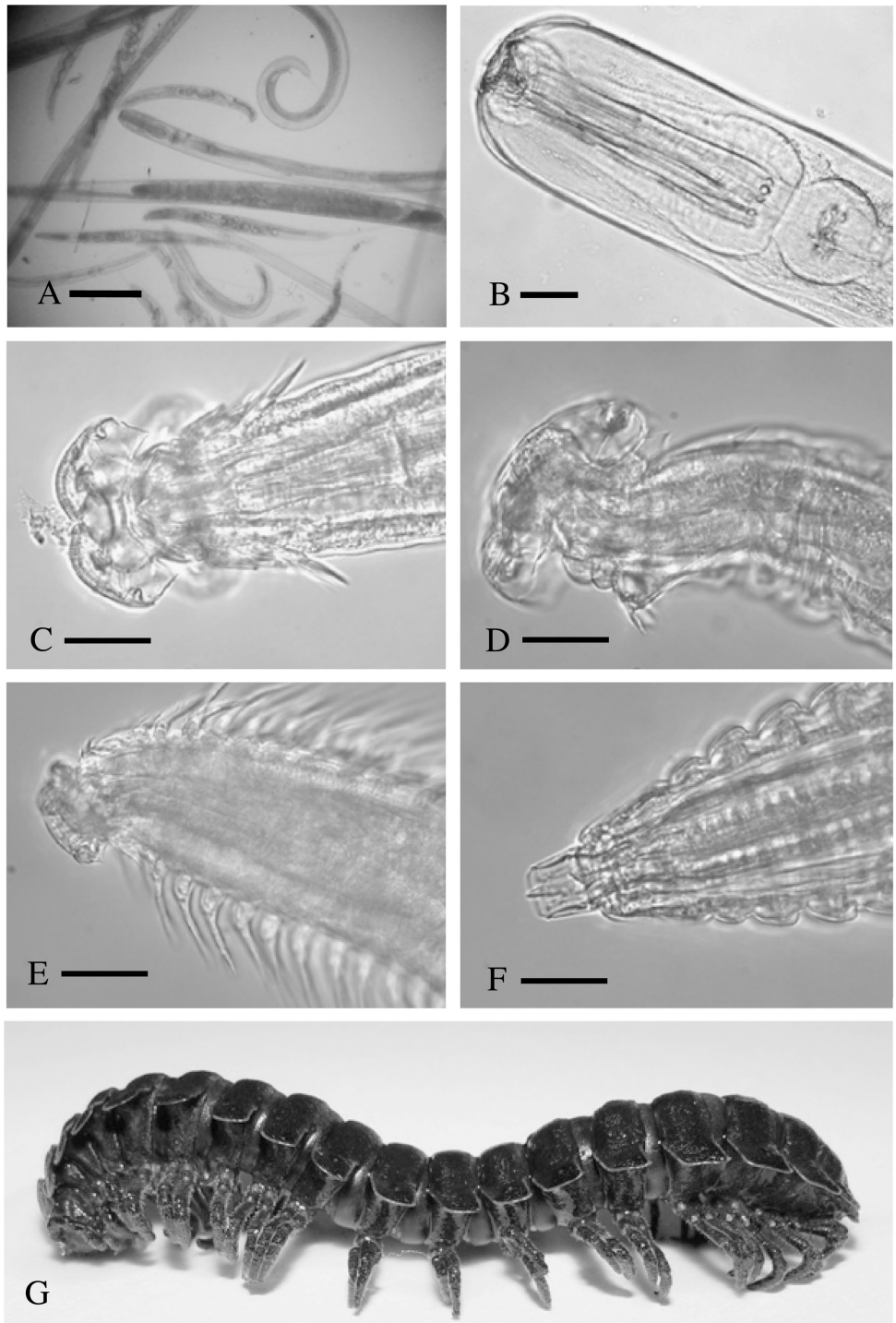


Fig. 3. *Nematodes recovered from Peruvian millipede preserved in DESS solution. A: Low magnification digital image of nematodes extracted directly from gut of preserved millipede; B: Anterior end of* Rhigonema *sp.; C: Anterior end of female* Heth *sp. 1; D: Anterior end of female* Heth *sp. 1; E: Anterior end of female* Heth *sp. 2, showing longitudinal spine rows; F: Anterior end of male* Heth *sp. 2, showing anterior rows of cuticular tubercles; G: Peruvian millipede after preservation in DESS solution and before dissection. (Scale bars: A = 370 μm; B = 50 μm; C-F = 10 μm.)*

the oxyurids, currently known only from the guts of large diplopods from tropical regions. PCR and sequencing of the specimens resulted in fungal rDNA sequences, presumably due to contamination with fungal amplicons that may derive from the millipede gut contents. It therefore remains unclear for the time being if DESS can be used to preserve and recover DNA of arthropod parasites inside host tissues. Presumably, better results can be obtained if the host is dissected immediately after killing, and the entire gut removed from the body and preserved separately in DESS. This would ensure much faster penetration of DESS through the gut and any nematodes inside it.

Scanning electron microscopy of nematodes preserved in both DESS solution and 5% formalin for 24 h show comparable results in terms of external morphology (Fig. 4). Slight differences in cuticular and amphidial condition were observed, most noticeably the lack of amphidial secretions in the DESS-preserved specimens allowing the fovea to be seen (Fig. 4D, F, arrowheads). When specimens were examined again 2 weeks after initial SEM preparation, collapse of structures was observed in some of the DESS-preserved nematodes (Fig. 4E). This is most likely due to the fact that DESS solution does not harden tissues as do fixatives such as formalin, thereby allowing structures to remain intact for long periods of time. Nematodes preserved in DESS or formalin for 6 weeks also showed comparable morphological results with some of the DESS-preserved specimens showing cuticular anomalies (Fig. 4F). Given the inconsistencies of SEM preparation, tests on a variety of parasitic and free-living nematodes from a wide range of terrestrial, freshwater, and marine habitats should be conducted and may give differing results. Also, the length of time exposed to DESS solution needs to be further examined in order to determine how long nematodes can remain in the solution and still be suitable for SEM.

Conclusions

DESS solution works quickly to inactivate enzymes that degrade DNA through the combined effects of a severe osmotic shock followed by rapid transportation of disodium EDTA and sodium chloride into tissues as enabled by DMSO. The solution preserves morphology with similar quality to formalin fixation and allows for a substantial reduction of chemical health hazards during the preservation and processing of samples. Although less volatile and less toxic than formalin, DESS can cause mild skin irritations and its properties facilitate transportation of other substances into the body, including more hazardous toxins. Gloves are therefore a wise precaution when working with DMSO and DESS solution, especially in an environment where more dangerous chemicals are present.

Unlike ethanol, DESS solution is not flammable, allowing for the safe transportation of samples without requiring special precautions or permissions. Because of its high salt concentration, DESS spills are rather messy and it is recommended to place samples into plastic containers for transportation to avoid potential breakage (Seutin *et al*., 1991). Safe and easy transportation of samples has become a major concern in recent years. The use of DESS solution greatly facilitates international shipment of important material while still allowing subsequent video recording and/or PCR analysis.

While it will take some years to establish the long-term effects on nematode morphology of preservation with DESS, we speculate that traditional refrigeration or freezing of specimens will substantially prolong the preservation of DNA compared to storage in DESS at room temperature. Because PCR amplifiable samples can be left in DESS at room temperature for months at a time, refrigerator and freezer space can be reserved for only the most important, temperature-sensitive material. Unlike fresh samples, DESS-preserved samples can be examined thoroughly and at length, without concern that interesting specimens will die before they can be picked out and prepared for PCR. Furthermore, DESS solution allows for the application of many traditional sample and specimen preparation methods for microscopy, in many cases retaining the ability to use these same nematodes for PCR. Based on this combination of features, we predict that during the next few years DESS solution will largely replace formalin for many nematological purposes, and especially where long-term preservation of morphology is not essential.

Acknowledgements

We would like to thank Drs Ion Toderas, Olga Kiseliova, Nadia Adreev, Sergiu Adreev, and Galina Busmschiu for their help in organising sampling trips and acting as guides in Moldova. We thank Dr Seymour Van Gundy for helping to organise the trip to Moldova, and Jennifer Haynes, Adler Dillman, and Elmer Parwani for their help in taking and extracting samples in Moldova. Also, we thank Dr Edward Platzer for providing initial references

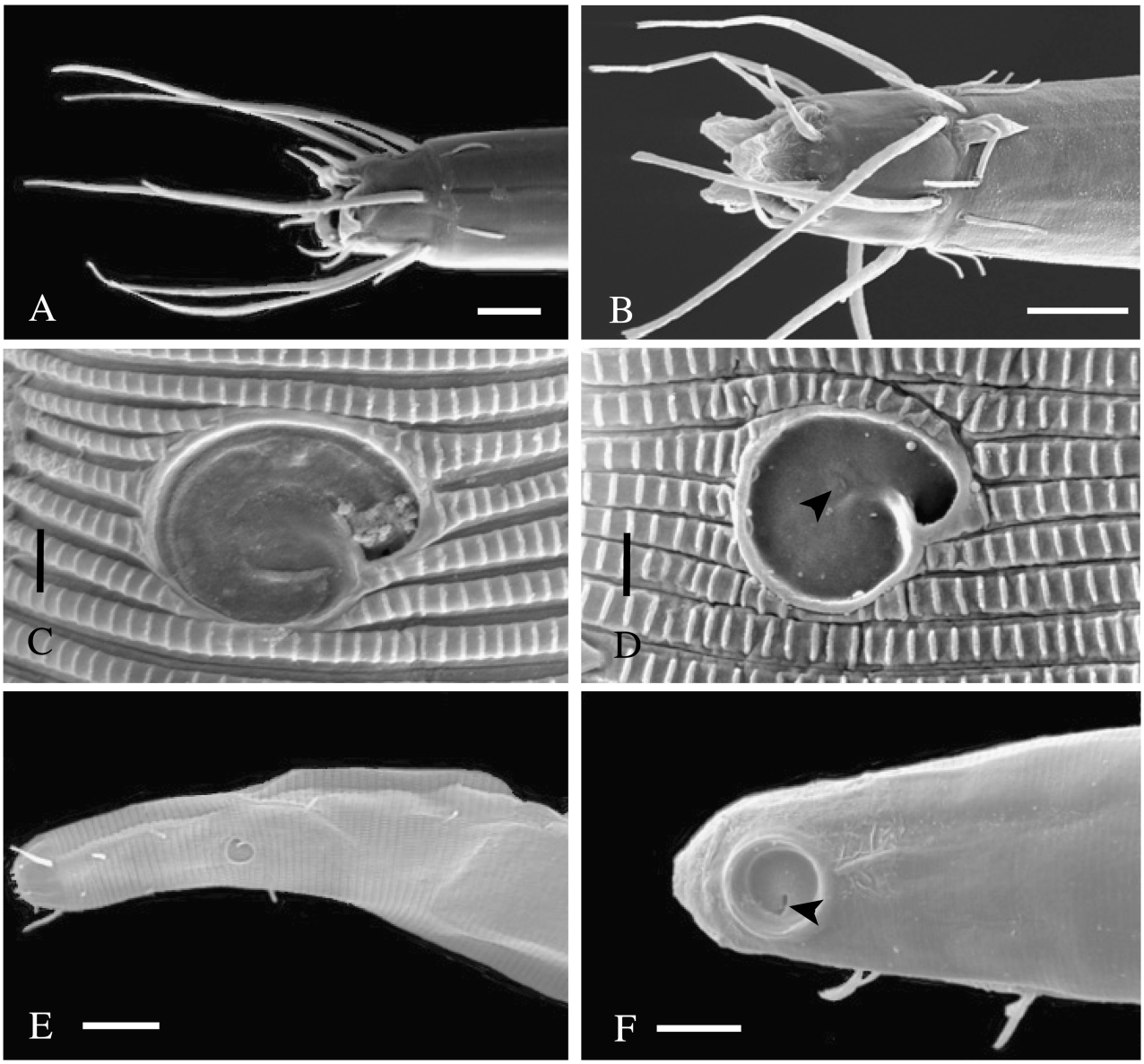


Fig. 4. *Nematodes prepared for Scanning Electron Microscopy (SEM). A: Anterior end of* Enoplolaimus*, preserved in formalin for 24 h before SEM preparation; B: Anterior end of* Enoplolaimus*, preserved in DESS solution for 24 h before SEM preparation; C:* Calomicrolaimus*, amphid, preserved in formalin for 24 h prior to SEM preparation; D:* Calomicrolaimus*, amphid, preserved in DESS solution for 24 h prior to SEM preparation, showing the amphidial fovea (arrowhead); E: Anterior end of* Calomicrolaimus*, preserved in DESS solution for 24 h prior to SEM preparation and 2 weeks after being prepared for SEM; F: Anterior end of* Terschellingia*, preserved in DESS solution for 6 weeks before being prepared for SEM, showing amphidial fovea (arrowhead). (Scale bars: A, E = 10 μm; B, F = 5 μm; C, D = 1 μm.)*

and guidance while in Moldova. We thank Dr Michael McKenry for supplying samples from Kern County, CA, USA, and Dr David Hunt for tentative identification of the millipede nematodes. We gratefully acknowledge support from the National Science Foundation through awards EF-0228692 and DEB-0315829, the National Research Council through award 001708, and the Central Science Laboratory, UK.

References

ALTSCHUL, S.F., MADDEN, T.L., SCHÄFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389-3402.

BLAXTER, M.L., DE LEY, P., GAREY, J.R., LIU, L.X., SCHELDEMAN, P., VIERSTRAETE, A., VANFLETEREN, J.R., MACKEY, L.Y., DORRIS, M., FRISSE, L.M., VIDA, J.T. & THOMAS, W.K. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71-75.

DAWSON, M.N., RASKOFF, K.A. & JACOBS, D.K. (1998). Field preservation of marine invertebrate tissue for DNA analyses. *Molecular Marine Biology and Biotechnology* 7, 145-152.

DE GRISSE, A.T. (1969). Redescription ou modifications de quelques techniques utilisées das l'étude des nématodes phytoparasitaires. *Mededeelingen Rijksfakulteit Landbouwwetenschappen Gent* 34, 351-369.

DE LEY, P. & BERT, W. (2002). Video capture and editing as a tool for the storage, distribution, and illustration of morphological characters of nematodes. *Journal of Nematology* 34, 269-302.

DE LEY, P. & BLAXTER, M. (2002). Systematic position and phylogeny. In: Lee, D.L. (Ed.). *The biology of nematodes*. London, UK, Taylor & Francis, pp. 1-30.

DE LEY, P., FELIX, M., FRISSE, L.M., NADLER, S.A., STERNBERG, P.W. & THOMAS, W.K. (1999). Molecular and morphological characterisation of two reproductively isolated species with mirror-image anatomy (Nematoda: Cephalobidae). *Nematology* 1, 591-612.

DESSAUER, H.C., COLE, C.J. & HAFNER, M.S. (1996). Collection and storage of tissues. In: Hillis, D.M., Moritz, C. & Mable, B.K. (Eds). *Molecular systematics, 2nd edn.* Sunderland, MA, USA, Sinauer Associates, pp. 29-47.

DILLON, N., AUSTIN, A.D. & BAROWSKY, E. (1996). Comparison of preservation techniques for DNA extraction from hymenopterous insects. *Insect Molecular Biology* 5, 21-24.

DIXON, M. & WEBB, E.C. (1979). *Enzymes.* New York, NY, USA, Academic Press, 1116 pp.

DORRIS, M., VINEY, M.E. & BLAXTER, M.L. (2002). Molecular phylogenetic analysis of the genus *Strongyloides* and related nematodes. *International Journal of Parasitology* 30, 1507-1517.

FLOYD, R., ABEBE, E., PAPERT, A. & BLAXTER, M.L. (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology* 11, 839-850.

GELFLAND, D.H. (1989). Taq DNA Polymerase. In: Erlich, H.A. (Ed.). *PCR technology: principles and applications for DNA amplification*. New York, NY, USA, Stockton Press, pp. 17-22.

GREER, C.E., PERERSON, S.L., KIVIAT, N.B. & MANOS, M.M. (1991). PCR amplification from paraffin-embedded tissues effects of fixative and fixation time. *American Journal of Clinical Pathology* 95, 117-124.

HOLZMANN, M. & PAWLOWSKI, J. (1996). Preservation of foraminifera for DNA extraction and PCR amplification. *Journal of Foraminiferal Research* 26, 264-267.

JACOBS, S.W. (1971). Pharmacology of DMSO. In: Jacobs, S.W., Rosenbaum, E.E. & Wood, D.C. (Eds). *Dimethyl sulfoxide*. New York, NY, USA, Marcel Dekker, Inc., pp. 99-112.

KILPATRICK, C.W. (2002). Noncryogenic preservation of mammalian tissues for DNA extraction: an assessment of storage methods. *Biochemical Genetics* 40, 53-62.

MILLER, K.S. & HOOK, L.A. (1996). Interim and archival preservation of plant specimens in alcohols for DNA studies. *Biotechniques* 20, 657-660.

RAMMLER, D.H. (1971). Use of DMSO in enzyme-catalyzed reactions. In: Jacobs, S.W., Rosenbaum, E.E. & Wood, D.C. (Eds). *Dimethyl sulfoxide*. New York, NY, USA, Marcel Dekker, Inc., pp. 189-206.

ROUBTSOVA, T.V., MOENS, M. & SUBBOTIN, S.A. (2005). PCR amplification of a rRNA gene fragment from formalin-fixed and glycerine-embedded nematodes from permanent slides. *Russian Journal of Nematology* 13, 137-140.

SCHANDER, C. & HALANYCH, K.M. (2003). DNA, PCR and formalinized animal tissue – short review and protocols. *Organisms, Diversity & Evolution* 3, 195-205.

SCHINDLER, A.F. (1961). A simple substitute for a Baermann funnel. *Plant Disease Reporter* 45, 747-748.

SEINHORST, J.W. (1959). A rapid method for the transfer of nematodes from fixative to anhydrous glycerin. *Nematologica* 4, 67-69.

SEUTIN, G., WHITE, B.N. & BOAG, P.T. (1991). Preservation of avian blood tissue samples for DNA analyses. *Canadian Journal of Zoology-Revue Canadienne de Zoologie* 69, 82-90.

TANDINGAN DE LEY, I., DE LEY, P., VIERSTRAETE, A., KARSSEN, G., MOENS, M. & VANFLETEREN, J. (2002). Phylogenetic analyses of *Meloidogyne* small subunit rDNA. *Journal of Nematology* 34, 319-327.

THOMAS, W.K., VIDA, J.T., FRISSE, L.M., MUNDO, M. & BALDWIN, J.G. (1997). DNA sequences from formalin-fixed nematodes: integrating molecular and morphological approaches to taxonomy. *Journal of Nematology* 29, 250-254.

TOE, L., BACK, C., ADJAMI, A.G., TANG, J.M. & UNNASCH, T.R. (1997). *Onchocerca volvulus*: comparison of field collection methods for the preservation of parasite and vector samples for PCR analysis. *Bulletin of the World Health Organization* 75, 443-447.

WILLIAMS, B.D., SCHRANK, B., HUYNH, C., SHOWNKEEN, R. & WATERSTON, R.H. (1992). A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* 131, 609-624.

Appendix 4.4

Tardigrade specimen identification including VCE number, relative size (small, medium or large), descriptions of cuticle (int = internal; ext = external), pharyngeal bulb morphology (including numbers and descriptions of macro- and microplacoids) and claws (a/spine = accessory spine; dbl = double; ext = external; int = internal; lgr = larger; shp = shape; w/ = with).

ID	VCE #	Morphological ID	Size	Cuticle	Pharyngeal bulb	Claws
HI14	383	Echinischus T1	sm	Plates, int/ext cirrus, cirrus A, cephalic papillae, red pigmentation	Circular, no placoids	Single claws, thickening as base of claws, hind legs w/ dentate collar
CA02	321	Isohypsibius TI	sm	No plates	Circular, 2 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of secondary arm)
CA41	360	Isohypsibius TI	sm	No plates	Circular, 2 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of secondary arm)
HI28	397	Isohypsibius TI	sm	No plates	Circular, 2 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of secondary arm)
CA22	341	Isohypsibius TI var I	sm	No plates	Circular, 4 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of secondary arm)
CA40	359	Isohypsibius TI var I	sm	No plates	Circular, 4 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of secondary arm)
CA45	364	Isohypsibius TI var I	sm	No plates	Circular, 4 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, difficult to make out detail
CA47	366	Isohypsibius TI var I	sm	No plates	Circular, 4 dorsal (1 widened, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of secondary arm)
CA32	351	Isohypsibius TI var II	med	No plates	Circular, 4 dorsal (1 elongated, 2), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of

						secondary arm)
D05_16	238	Isohypsibius TII	med	No plates	Oval, difficult to see	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a approx third-half of secondary arm)
FD01	316	Macrobiotus TI	med	No plates	Oval, 2 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
FD03	318	Macrobiotus TI	med	No plates	Oval, 2 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
FD04	319	Macrobiotus TI	med	No plates	Oval, 2 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
HI01	370	Macrobiotus TI	med	No plates	Oval, 2 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
HI31	400	Macrobiotus TI	med	No plates	Oval, 2 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI08	377	Macrobiotus TI ?	med	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
HI17	386	Macrobiotus TI var I	med	No plates	Circular, 4 dorsal (1 elongated, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
HI19	388	Macrobiotus TI var I	med	No plates	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
FD02	317	Macrobiotus TI var I	med	No plates	Oval, 2 dorsal 2 ventral, (1 elongated, 2) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, deep v shp, spur at base, lunule w/out teeth
CA27	346	Macrobiotus TII	med	No plates	Oval, 4 dorsal (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, v shpd, spur at base, lunule w/out teeth
CA01	320	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth

CA03	322	Macrobiotus TII	lg	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA07	326	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA08	327	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA09	328	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA10	329	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA11	330	Macrobiotus TII	sm	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA13	332	Macrobiotus TII	lg	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA16	335	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA19	338	Macrobiotus TII	sm	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA20	339	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA21	340	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA23	342	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth

CA31	350	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA33	352	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA34	353	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA35	354	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA39	358	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI24	393	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI46	415	Macrobiotus TII	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA06	325	Macrobiotus TII var I	sm	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA43	362	Macrobiotus TII var I	med	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA15	334	Macrobiotus TII var II	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, ext almost at right-angle, v shp, spur at base, lunule w/out teeth
CA18	337	Macrobiotus TII var II	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, ext almost at right-angle, v shp, spur at base, lunule w/out teeth
CA37	356	Macrobiotus TII var II	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, ext almost at right-angle, v shp, spur at base, lunule w/out teeth
HI38	407	Macrobiotus TII var III	sm	No plates	Circular, 2 dorsal 2 ventral (1, 2, 3), no visible microplacoids	Dbl claw, unsymmetrical, int symmetrical, ext w/ long post branch (intersection at a third of

						secondary arm)
HI43	412	Macrobiotus TII var III	sm	No plates	Circular, 2 dorsal 2 ventral (1, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI44	413	Macrobiotus TII var III	med	No plates	Circular, 4 dorsal (1, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI40	409	Macrobiotus TII var III	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI45	414	Macrobiotus TII var III	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI39	408	Macrobiotus TII var III	med	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI36	405	Macrobiotus TII var III	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI32	401	Macrobiotus TII var IV	med	No plates	Circular, 2 dorsal 2 ventral (1, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA28	347	Macrobiotus TII var V	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA36	355	Macrobiotus TII var V	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA48	367	Macrobiotus TII var V	med	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA46	365	Macrobiotus TII var V	sm	No plates	Oval, macroplacoids present but difficult to count	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA25	344	Macrobiotus TII var VI	med	No plates	Circular, 4 dorsal (1, 2, 3 elongated), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA05	324	Macrobiotus TIII	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA29	348	Macrobiotus TIII	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA38	357	Macrobiotus TIII	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth

CA44	363	Macrobiotus TIII	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA04	323	Macrobiotus TIII	med	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA26	345	Macrobiotus TIII	med	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
HI05	374	Macrobiotus TIII	med	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
CA49	368	Macrobiotus TIII var I	sm	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI06	375	Macrobiotus TIII var I	sm	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI27	396	Macrobiotus TIII var I	sm	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA17	336	Macrobiotus TIII var I	lg	No plates	No visible bulb	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
CA12	331	Macrobiotus TIII var I	med	No plates	Circular, no placoids	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
HI10	379	Macrobiotus TIV (cf M. occidentalis)	med	No plates, red pigmentation, almond pores	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
HI22	391	Macrobiotus TIV (cf M. occidentalis)	med	No plates, red pigmentation, almond pores	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
HI23	392	Macrobiotus TIV (cf M. occidentalis)	med	No plates, red pigmentation, almond pores	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
HI26	395	Macrobiotus TIV (cf M. occidentalis)	med	No plates, red pigmentation, almond pores	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
HI41	410	Macrobiotus TIV var I	sm	No plates, almond pores	Circular, 4 dorsal (1, 2), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w a/spine, Y shpd, spur at base, lunule w/out teeth
HI29	398	Macrobiotus TV	sm	No plates	Circular, 2 dorsal 2 ventral (1, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, v shp, spur at base, lunule present - sm
HI13	382	Macrobiotus TV	sm	No plates	Circular, 4 dorsal (1, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, v shp, spur at base, lunule present - sm

HI20	389	Macrobiotus TV	sm	No plates	Circular, 4 dorsal (1, 2, 3), no visible microplacoids	Dbl claw, roughly symmetrical, lgr int branch, v shp, spur at base, lunule present - sm
CA24	343	Macrobiotus TV	sm	No plates	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, v shp, spur at base, lunule present - sm
CA42	361	Macrobiotus TV	sm	No plates	Circular, 4 dorsal 2 ventral (1, 2, 3 elongated), microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, v shp, spur at base, lunule present - sm
HI25	394	Macrobiotus TV var I	sm	No plates	Circular, macroplacoids present but difficult to count	Dbl claw, roughly symmetrical, lgr int branch, v shp, spur at base, lunule present - sm
D12_03	273	Macrobiotus TV var II	sm	No plates, eye spots	Oval, 4 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
D12_04	274	Macrobiotus TV var II	sm	No plates, eye spots	Oval, 4 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
D12_06	276	Macrobiotus TV var III	sm	No plates, sm pores?	Oval, 4 dorsal 2 ventral (1 elongated, 2) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, lgr int branch, w/ a/spine, v shp, spur at base, lunule w/out teeth
D12_05	275	Macrobiotus TV var III	sm	No plates	Oval, 4 dorsal 2 ventral (1, 2, 3) macroplacoids, microplacoids present	Dbl claw, roughly symmetrical, v shp, difficult to see
CA14	333	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI02	371	Milnesium TI	lg	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI03	372	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI04	373	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI07	376	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer

						than 4th
HI09	378	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI12	381	Milnesium TI	sm	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI16	385	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI18	387	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI21	390	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI33	402	Milnesium TI	sm	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI34	403	Milnesium TI	sm	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI35	404	Milnesium TI	med	No plates	Pear shpd, no placoids, papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI11	380	Milnesium TI var I	med	No plates	Pear shpd, no placoids, no visible papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI15	384	Milnesium TI var I	sm	No plates	Pear shpd, no placoids, no visible papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI30	399	Milnesium TI var I	med	No plates	Pear shpd, no placoids, no visible papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer than 4th
HI37	406	Milnesium TI var I	sm	No plates	Pear shpd, no placoids, no visible papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ trifurcate secondary arm, first 3 legs longer

						than 4th
HI42	411	Milnesium TI var I	med	No plates	Pear shpd, no placoids, no visible papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm, difficult to see
D05_01	223	Milnesium TI var I	med	No plates	Pear shpd, no placoids, no visible papillae around mouth	Dbl claw, unsymmetrical, elongated ant principle arm w/ bifurcate secondary arm?

Appendix 4.5

Details of resample MOTUs at plateau phases from the tardigrade survey. COI plateau was from 9 – 60 bp, but there was very little change so only two cut-offs are displayed. Number and member sequences (D = Disko Island; FD = Florida; CA = California; HI = Hawaii) of primary runs are listed. Classification of resample MOTUs compared to primary runs (E; S; J; C) as in Appendix 4.3.

COI	#	Members	E	S	J	C	Total
9bp_MOTU12	1	D05_01	100	0	0	0	100
9bp_MOTU11	1	D05_16	100	0	0	0	100
9bp_MOTU10	1	D12_06	100	0	0	0	100
9bp_MOTU09	5	HI23 HI41 HI22 HI26 HI10	100	0	0	0	100
9bp_MOTU08	2	HI38 HI28	100	0	0	0	100
9bp_MOTU07	2	D12_03 D12_04	100	0	0	0	100
9bp_MOTU06	1	HI44	100	0	0	0	100
9bp_MOTU05	1	HI12	100	0	0	0	100
9bp_MOTU04	1	CA32	100	0	0	0	100
9bp_MOTU03	10	HI04 HI18 HI09 HI33 HI21 HI35 HI03 HI34 HI30 HI02	99	1	0	0	100
9bp_MOTU02	17	HI40 HI01 HI36 HI20 HI39 HI27 HI25 HI24 HI19 HI08 HI05 HI43 HI31 HI17 HI13 HI45 HI29	100	0	0	0	100
9bp_MOTU01	39	CA21 CA10 CA12 CA49 CA17 CA44 CA34 CA38 CA31 CA36 CA35 CA03 CA07 CA46 CA11 CA01 CA06 CA37 CA26 CA08 CA43 CA29 CA05 CA24 CA27 CA04 CA09 CA48 CA39 CA15 CA33 CA23 CA42 CA18 CA19 CA28 CA16 CA20 CA25	99	1	0	0	100
9bp_MOTU	12		1198	2	0	0	1200
10bp_MOTU12	5	HI22 HI10 HI41 HI23 HI26	100	0	0	0	100
10bp_MOTU11	1	D05_01	100	0	0	0	100
10bp_MOTU10	2	D12_04 D12_03	100	0	0	0	100
10bp_MOTU09	1	HI44	100	0	0	0	100
10bp_MOTU08	1	D05_16	100	0	0	0	100
10bp_MOTU07	1	HI12	100	0	0	0	100
10bp_MOTU06	1	CA32	100	0	0	0	100
10bp_MOTU05	1	D12_06	100	0	0	0	100
10bp_MOTU04	10	HI02 HI30 HI09 HI04 HI33 HI35 HI03 HI21 HI18 HI34	100	0	0	0	100
10bp_MOTU03	2	HI38 HI28	100	0	0	0	100
10bp_MOTU02	17	HI31 HI05 HI19 HI40 HI08 HI29 HI13 HI20 HI45 HI24 HI27 HI39 HI01 HI17 HI25 HI43 HI36	100	0	0	0	100
10bp_MOTU01	39	CA44 CA07 CA33 CA05 CA37 CA01 CA03 CA46 CA35 CA42 CA34 CA25 CA08 CA43 CA24 CA48 CA18 CA38 CA15 CA39 CA17 CA11 CA26 CA12 CA23 CA20 CA49 CA06 CA21 CA31 CA19 CA09 CA29 CA28 CA27 CA16 CA04 CA36 CA10	100	0	0	0	100
10bp_MOTU	12		1200	0	0	0	1200
LSU	**#**	**Members**	**E**	**S**	**J**	**C**	**Total**
16bp_MOTU15	1	HI06	99	0	1	0	100
16bp_MOTU14	1	CA32	100	0	0	0	100
16bp_MOTU13	1	D12_05	100	0	0	0	100

16bp_MOTU12	2	HI32 HI44	100	0	0	0	100
16bp_MOTU11	2	D12_03 D12_04	84	0	16	0	100
16bp_MOTU10	3	FD01 FD03 FD04	100	0	0	0	100
16bp_MOTU09	18	HI39 HI20 HI05 HI45 HI43 HI31 HI24 HI01 HI17 HI25 HI19 HI36 HI40 HI27 HI29 HI08 HI13 HI46	80	0	20	0	100
16bp_MOTU08	1	D12_06	96	0	4	0	100
16bp_MOTU07	1	HI28	100	0	0	0	100
16bp_MOTU06	3	CA45 CA47 CA02	100	0	0	0	100
16bp_MOTU05	2	CA22 CA40	100	0	0	0	100
16bp_MOTU04	5	HI23 HI41 HI10 HI22 HI26	90	0	10	0	100
16bp_MOTU03	1	CA41	100	0	0	0	100
16bp_MOTU02	39	CA27 CA08 CA37 CA06 CA36 CA46 CA44 CA01 CA20 CA21 CA12 CA29 CA28 CA23 CA16 CA26 CA10 CA05 CA17 CA38 CA03 CA34 CA25 CA04 CA11 CA31 CA18 CA33 CA24 CA49 CA13 CA39 CA09 CA19 CA35 CA43 CA07 CA42 CA48	79	0	21	0	100
16bp_MOTU01	18	HI11 HI02 HI03 HI18 HI33 CA14 HI42 HI21 HI04 HI35 HI09 HI15 HI30 HI37 HI07 HI16 HI34 HI12	29	0	71	0	100
16bp_MOTU	15		1357	0	143	0	1500
17bp_MOTU14	1	HI06	100	0	0	0	100
17bp_MOTU13	2	D12_03 D12_04	86	0	14	0	100
17bp_MOTU12	1	D12_05	100	0	0	0	100
17bp_MOTU11	1	CA32	100	0	0	0	100
17bp_MOTU10	2	CA22 CA40	100	0	0	0	100
17bp_MOTU09	2	HI32 HI44	100	0	0	0	100
17bp_MOTU08	1	CA41	100	0	0	0	100
17bp_MOTU07	1	D12_06	99	0	1	0	100
17bp_MOTU06	1	HI28	100	0	0	0	100
17bp_MOTU05	3	CA45 CA47 CA02	100	0	0	0	100
17bp_MOTU04	5	HI10 HI22 HI41 HI23 HI26	98	0	2	0	100
17bp_MOTU03	39	CA16 CA25 CA46 CA18 CA12 CA39 CA37 CA26 CA06 CA05 CA35 CA11 CA33 CA27 CA10 CA04 CA24 CA49 CA19 CA20 CA42 CA36 CA08 CA23 CA13 CA34 CA29 CA28 CA03 CA38 CA09 CA17 CA07 CA01 CA44 CA48 CA43 CA21 CA31	68	0	32	0	100
17bp_MOTU02	36	HI45 HI37 HI13 HI09 HI04 HI03 HI19 HI33 HI16 HI43 HI39 HI24 HI25 HI02 HI18 CA14 HI07 HI29 HI21 HI17 HI01 HI05 HI27 HI20 HI15 HI31 HI40 HI12 HI46 HI11 HI30 HI08 HI42 HI36 HI35 HI34	15	36	1	48	100
17bp_MOTU01	3	FD04 FD01 FD03	100	0	0	0	100
17bp_MOTU	14		1266	36	50	48	1400
18bp_MOTU14	1	D12_06	96	0	4	0	100
18bp_MOTU13	1	CA41	100	0	0	0	100
18bp_MOTU12	2	CA40 CA22	100	0	0	0	100
18bp_MOTU11	1	CA32	100	0	0	0	100
18bp_MOTU10	3	FD01 FD03 FD04	100	0	0	0	100
18bp_MOTU09	1	D12_05	100	0	0	0	100
18bp_MOTU08	5	HI10 HI26 HI41 HI22 HI23	89	0	11	0	100
18bp_MOTU07	3	CA47 CA45 CA02	100	0	0	0	100
18bp_MOTU06	2	D12_04 D12_03	91	0	9	0	100
18bp_MOTU05	1	HI28	100	0	0	0	100
18bp_MOTU04	2	HI44 HI32	97	0	3	0	100
18bp_MOTU03	1	HI06	97	0	3	0	100

18bp_MOTU02	18	HI29 HI05 HI46 HI40 HI24 HI08 HI25 HI31 HI19 HI39 HI27 HI13 HI01 HI36 HI43 HI17 HI20 HI45	86	0	14	0	100
18bp_MOTU01	57	HI18 CA33 HI33 CA48 CA19 CA03 CA25 HI04 CA23 HI34 HI16 HI02 CA14 HI42 CA28 CA37 CA07 CA20 CA26 CA21 CA29 CA38 HI21 CA39 CA11 HI12 CA12 CA06 CA18 CA05 HI37 HI11 CA35 CA17 HI15 CA43 CA08 CA09 CA04 CA42 CA27 CA49 CA10 CA46 CA13 CA44 HI09 CA24 CA01 HI03 CA36 CA34 HI07 HI35 CA31 CA16 HI30	41	27	2	30	100
18bp_MOTU	14		1297	27	46	30	1400
19bp_MOTU14	2	D12_04 D12_03	94	0	6	0	100
19bp_MOTU13	1	D12_05	100	0	0	0	100
19bp_MOTU12	1	CA41	100	0	0	0	100
19bp_MOTU11	1	HI06	100	0	0	0	100
19bp_MOTU10	1	CA32	100	0	0	0	100
19bp_MOTU09	2	HI44 HI32	99	0	1	0	100
19bp_MOTU08	1	D12_06	95	0	5	0	100
19bp_MOTU07	2	CA40 CA22	100	0	0	0	100
19bp_MOTU06	3	CA02 CA47 CA45	100	0	0	0	100
19bp_MOTU05	18	HI40 HI13 HI39 HI20 HI45 HI17 HI31 HI24 HI01 HI08 HI05 HI46 HI27 HI19 HI36 HI25 HI43 HI29	84	0	16	0	100
19bp_MOTU04	1	HI28	100	0	0	0	100
19bp_MOTU03	23	HI41 HI42 HI15 HI18 HI04 HI07 HI02 HI21 HI16 HI30 HI22 HI33 CA14 HI26 HI23 HI35 HI10 HI34 HI11 HI37 HI03 HI12 HI09	11	7	0	82	100
19bp_MOTU02	39	CA36 CA42 CA08 CA29 CA12 CA44 CA16 CA09 CA11 CA20 CA23 CA46 CA35 CA24 CA10 CA49 CA26 CA21 CA05 CA04 CA03 CA18 CA33 CA39 CA17 CA07 CA27 CA31 CA25 CA06 CA43 CA19 CA34 CA37 CA28 CA38 CA48 CA01 CA13	43	0	57	0	100
19bp_MOTU01	3	FD01 FD03 FD04	100	0	0	0	100
19bp_MOTU	14		1226	7	85	82	1400
20bp_MOTU14	1	D12_06	98	0	2	0	100
20bp_MOTU13	1	CA32	100	0	0	0	100
20bp_MOTU12	1	HI28	100	0	0	0	100
20bp_MOTU11	2	D12_03 D12_04	87	0	13	0	100
20bp_MOTU10	2	CA40 CA22	100	0	0	0	100
20bp_MOTU09	3	FD01 FD04 FD03	100	0	0	0	100
20bp_MOTU08	1	CA41	100	0	0	0	100
20bp_MOTU07	1	HI06	100	0	0	0	100
20bp_MOTU06	2	HI32 HI44	100	0	0	0	100
20bp_MOTU05	3	CA02 CA47 CA45	100	0	0	0	100
20bp_MOTU04	5	HI22 HI26 HI23 HI41 HI10	90	0	10	0	100
20bp_MOTU03	18	HI36 HI39 HI25 HI43 HI20 HI19 HI08 HI31 HI24 HI13 HI46 HI40 HI45 HI01 HI05 HI17 HI27 HI29	79	0	21	0	100
20bp_MOTU02	1	D12_05	100	0	0	0	100
20bp_MOTU01	57	HI16 HI11 CA28 CA11 CA04 CA42 CA39 HI21 CA49 CA27 CA16 CA21 CA25 CA26 CA23 CA07 HI37 CA06 CA29 CA34 CA09 CA03 CA24 CA36 CA19 HI34 CA37 CA10 CA18 HI03 CA14 HI42 HI33 CA33 CA17 HI15 CA05 HI04 CA48 HI09 HI18 CA44 HI30 CA13 HI12 CA35 CA01 CA43 CA38 HI07 HI35 CA31 CA46 CA08 HI02 CA12 CA20	52	2	8	38	100
20bp_MOTU	14		1306	2	54	38	1400

SSU	#	Members	E	S	J	C	Total
7bp_MOTU11	1	CA32	100	0	0	0	100
7bp_MOTU10	1	D12_06	100	0	0	0	100
7bp_MOTU09	1	D05_16	100	0	0	0	100
7bp_MOTU08	2	HI28 HI38	70	30	0	0	100
7bp_MOTU07	3	HI06 D12_04 D12_03	100	0	0	0	100
7bp_MOTU06	1	D05_01	100	0	0	0	100
7bp_MOTU05	7	HI22 HI10 HI23 HI44 HI32 HI26 HI41	96	4	0	0	100
7bp_MOTU04	6	CA45 CA02 CA47 CA41 CA22 CA40	100	0	0	0	100
7bp_MOTU03	18	HI21 HI15 HI11 HI42 HI12 HI37 HI34 HI35 HI03 HI09 HI33 HI18 CA14 HI04 HI30 HI02 HI16 HI07	100	0	0	0	100
7bp_MOTU02	37	CA44 CA03 CA04 CA46 CA16 CA20 CA25 CA34 CA31 CA07 CA17 CA36 CA19 CA12 CA38 CA35 CA28 CA23 CA43 CA09 CA49 CA11 CA13 CA26 CA06 CA05 CA39 CA27 CA48 CA29 CA10 CA42 CA01 CA18 CA21 CA24 CA37	100	0	0	0	100
7bp_MOTU01	21	D12_05 HI39 HI43 HI19 HI05 FD01 HI08 HI13 HI40 HI24 HI45 HI31 HI46 HI20 HI27 HI25 HI01 HI29 FD03 HI36 HI17	100	0	0	0	100
7bp	11		1066	34	0	0	1100
8bp_MOTU12	1	HI38	33	0	67	0	100
8bp_MOTU11	1	HI28	33	0	67	0	100
8bp_MOTU10	1	D05_16	100	0	0	0	100
8bp_MOTU09	1	CA32	100	0	0	0	100
8bp_MOTU08	1	D12_06	100	0	0	0	100
8bp_MOTU07	1	D05_01	100	0	0	0	100
8bp_MOTU06	18	HI03 HI42 HI35 CA14 HI11 HI15 HI16 HI07 HI30 HI12 HI37 HI04 HI21 HI18 HI02 HI34 HI09 HI33	100	0	0	0	100
8bp_MOTU05	3	D12_04 HI06 D12_03	100	0	0	0	100
8bp_MOTU04	6	CA40 CA47 CA45 CA02 CA41 CA22	100	0	0	0	100
8bp_MOTU03	37	CA06 CA29 CA01 CA35 CA37 CA13 CA49 CA04 CA10 CA39 CA42 CA28 CA26 CA46 CA07 CA27 CA48 CA09 CA21 CA12 CA05 CA19 CA44 CA38 CA18 CA03 CA20 CA11 CA23 CA17 CA43 CA36 CA16 CA34 CA25 CA24 CA31	100	0	0	0	100
8bp_MOTU02	21	HI39 HI13 HI24 HI25 FD01 HI17 HI43 HI05 HI19 HI40 HI20 HI45 HI27 HI29 HI01 HI31 HI36 FD03 HI46 HI08 D12_05	94	6	0	0	100
8bp_MOTU01	7	HI32 HI23 HI26 HI10 HI44 HI41 HI22	100	0	0	0	100
8bp	12		1060	6	134	0	1200
9bp_MOTU11	1	D05_16	100	0	0	0	100
9bp_MOTU10	1	D12_06	100	0	0	0	100
9bp_MOTU09	2	HI38 HI28	100	0	0	0	100
9bp_MOTU08	1	CA32	100	0	0	0	100
9bp_MOTU07	7	HI23 HI32 HI26 HI10 HI41 HI44 HI22	100	0	0	0	100
9bp_MOTU06	6	CA45 CA02 CA22 CA41 CA40 CA47	100	0	0	0	100
9bp_MOTU05	18	HI18 HI35 HI42 HI12 HI34 HI02 CA14 HI15 HI07 HI37 HI21 HI04 HI03 HI09 HI33 HI11 HI30 HI16	100	0	0	0	100
9bp_MOTU04	3	D12_04 HI06 D12_03	100	0	0	0	100
9bp_MOTU03	1	D05_01	100	0	0	0	100
9bp_MOTU02	21	HI17 HI19 HI31 HI24 HI25 D12_05 HI45 HI43 FD01 HI08 HI01 HI05 HI39 HI29 HI20 HI36 HI46 FD03 HI13 HI40 HI27	100	0	0	0	100
9bp_MOTU01	37	CA37 CA27 CA03 CA01 CA25 CA48 CA20 CA11	100	0	0	0	100

		CA09 CA05 CA39 CA49 CA23 CA04 CA19 CA36 CA35 CA28 CA31 CA34 CA21 CA46 CA12 CA38 CA29 CA13 CA10 CA16 CA07 CA42 CA06 CA18 CA26 CA44 CA17 CA24 CA43				
9bp	11		1100	0	0	0	1100
10bp_MOTU11	1	D12_06	100	0	0	0	100
10bp_MOTU10	1	CA32	100	0	0	0	100
10bp_MOTU09	1	D05_16	100	0	0	0	100
10bp_MOTU08	1	D05_01	100	0	0	0	100
10bp_MOTU07	2	HI28 HI38	57	43	0	0	100
10bp_MOTU06	7	HI23 HI41 HI22 HI44 HI32 HI26 HI10	100	0	0	0	100
10bp_MOTU05	3	HI06 D12_03 D12_04	100	0	0	0	100
10bp_MOTU04	6	CA47 CA40 CA41 CA02 CA45 CA22	100	0	0	0	100
10bp_MOTU03	21	HI05 HI29 HI20 HI43 HI17 HI08 HI27 HI46 HI40 FD01 FD03 HI19 HI39 D12_05 HI01 HI24 HI45 HI13 HI31 HI25 HI36	93	7	0	0	100
10bp_MOTU02	37	CA38 CA13 CA04 CA01 CA11 CA25 CA43 CA34 CA07 CA23 CA48 CA03 CA12 CA49 CA44 CA05 CA06 CA39 CA24 CA16 CA37 CA19 CA28 CA42 CA35 CA18 CA26 CA36 CA31 CA46 CA21 CA09 CA10 CA17 CA29 CA20 CA27	100	0	0	0	100
10bp_MOTU01	18	HI30 HI04 HI02 CA14 HI12 HI09 HI18 HI34 HI35 HI37 HI33 HI07 HI15 HI16 HI42 HI11 HI03 HI21	100	0	0	0	100
10bp	11		1050	50	0	0	1100
11bp_MOTU11	1	D05_01	100	0	0	0	100
11bp_MOTU10	1	CA32	100	0	0	0	100
11bp_MOTU09	1	D12_06	100	0	0	0	100
11bp_MOTU08	1	D05_16	100	0	0	0	100
11bp_MOTU07	2	HI38 HI28	66	34	0	0	100
11bp_MOTU06	3	D12_04 D12_03 HI06	100	0	0	0	100
11bp_MOTU05	21	HI19 D12_05 HI01 HI20 HI36 FD01 HI40 HI25 HI27 HI46 HI45 HI29 HI31 FD03 HI24 HI05 HI08 HI13 HI17 HI43 HI39	100	0	0	0	100
11bp_MOTU04	37	CA07 CA10 CA13 CA24 CA05 CA43 CA27 CA49 CA28 CA35 CA42 CA21 CA44 CA36 CA04 CA17 CA18 CA34 CA12 CA48 CA31 CA29 CA20 CA37 CA01 CA39 CA16 CA06 CA09 CA26 CA46 CA03 CA23 CA25 CA38 CA19 CA11	100	0	0	0	100
11bp_MOTU03	6	CA45 CA02 CA41 CA40 CA22 CA47	100	0	0	0	100
11bp_MOTU02	7	HI44 HI32 HI10 HI41 HI22 HI23 HI26	100	0	0	0	100
11bp_MOTU01	18	HI02 HI33 HI18 HI04 HI07 HI11 HI37 HI15 HI21 HI12 HI09 HI34 HI16 CA14 HI03 HI30 HI42 HI35	100	0	0	0	100
11bp	11		1066	34	0	0	1100
12bp_MOTU11	1	D12_06	100	0	0	0	100
12bp_MOTU10	1	D05_16	100	0	0	0	100
12bp_MOTU09	1	D05_01	100	0	0	0	100
12bp_MOTU08	2	HI38 HI28	71	29	0	0	100
12bp_MOTU07	3	D12_04 D12_03 HI06	66	0	34	0	100
12bp_MOTU06	1	CA32	100	0	0	0	100
12bp_MOTU05	6	CA41 CA02 CA47 CA22 CA45 CA40	100	0	0	0	100
12bp_MOTU04	7	HI44 HI32 HI22 HI10 HI41 HI23 HI26	66	0	34	0	100
12bp_MOTU03	18	HI16 HI07 HI34 HI37 HI03 HI18 CA14 HI02 HI04 HI33 HI09 HI30 HI12 HI35 HI42 HI11 HI15 HI21	100	0	0	0	100

12bp_MOTU02	37	CA07 CA25 CA35 CA42 CA16 CA17 CA13 CA43 CA24 CA20 CA44 CA36 CA39 CA34 CA06 CA11 CA10 CA04 CA38 CA49 CA37 CA09 CA46 CA01 CA18 CA12 CA21 CA19 CA29 CA23 CA48 CA27 CA05 CA03 CA26 CA28 CA31	100	0	0	0	100
12bp_MOTU01	21	HI39 HI01 HI25 HI19 FD03 HI29 HI13 HI36 HI43 HI45 HI27 HI17 HI31 HI46 FD01 HI24 HI05 HI40 HI20 HI08 D12_05	97	3	0	0	100
12bp	11		1000	32	68	0	1100
13bp_MOTU10	1	D12_06	100	0	0	0	100
13bp_MOTU09	1	CA32	100	0	0	0	100
13bp_MOTU08	1	D05_01	100	0	0	0	100
13bp_MOTU07	1	D05_16	100	0	0	0	100
13bp_MOTU06	2	HI38 HI28	68	32	0	0	100
13bp_MOTU05	10	HI41 HI06 HI32 D12_04 HI10 HI23 HI44 HI26 HI22 D12_03	34	66	0	0	100
13bp_MOTU04	18	HI37 HI04 HI18 HI42 HI11 HI21 HI12 HI15 HI33 HI03 CA14 HI09 HI35 HI34 HI16 HI30 HI07 HI02	100	0	0	0	100
13bp_MOTU03	37	CA17 CA09 CA42 CA19 CA12 CA26 CA36 CA23 CA21 CA39 CA43 CA18 CA03 CA35 CA48 CA04 CA05 CA37 CA01 CA38 CA24 CA16 CA46 CA49 CA29 CA11 CA28 CA13 CA07 CA27 CA06 CA25 CA20 CA10 CA34 CA31 CA44	100	0	0	0	100
13bp_MOTU02	6	CA45 CA41 CA02 CA40 CA47 CA22	100	0	0	0	100
13bp_MOTU01	21	HI43 HI20 HI01 HI27 HI45 HI36 HI17 HI08 HI29 HI46 HI39 HI19 HI25 D12_05 HI24 FD01 HI40 HI05 FD03 HI31 HI13	99	1	0	0	100
13bp	10		901	99	0	0	1000
14bp_MOTU12	1	D05_01	100	0	0	0	100
14bp_MOTU11	1	HI38	33	0	67	0	100
14bp_MOTU10	1	CA32	100	0	0	0	100
14bp_MOTU09	1	HI28	33	0	67	0	100
14bp_MOTU08	1	D05_16	100	0	0	0	100
14bp_MOTU07	7	HI23 HI32 HI10 HI26 HI22 HI41 HI44	69	0	31	0	100
14bp_MOTU06	6	CA41 CA22 CA02 CA45 CA40 CA47	100	0	0	0	100
14bp_MOTU05	21	HI45 HI31 HI01 HI17 HI36 HI08 HI13 HI25 HI46 HI43 HI40 D12_05 HI05 HI20 FD01 HI27 HI19 FD03 HI29 HI39 HI24	93	7	0	0	100
14bp_MOTU04	3	D12_03 HI06 D12_04	69	0	31	0	100
14bp_MOTU03	1	D12_06	100	0	0	0	100
14bp_MOTU02	18	HI04 HI03 HI21 HI15 HI16 HI42 HI07 HI34 HI37 HI30 HI11 HI09 HI35 HI12 HI33 HI18 CA14 HI02	100	0	0	0	100
14bp_MOTU01	37	CA06 CA24 CA11 CA12 CA23 CA48 CA44 CA38 CA19 CA03 CA18 CA49 CA25 CA13 CA31 CA37 CA29 CA26 CA34 CA42 CA10 CA07 CA28 CA46 CA27 CA39 CA01 CA04 CA43 CA35 CA16 CA05 CA17 CA09 CA36 CA20 CA21	100	0	0	0	100
14bp	12		997	7	196	0	1200
15bp_MOTU11	1	D05_16	100	0	0	0	100
15bp_MOTU10	1	HI38	33	0	67	0	100
15bp_MOTU09	1	D05_01	100	0	0	0	100
15bp_MOTU08	1	CA32	100	0	0	0	100
15bp_MOTU07	1	D12_06	100	0	0	0	100
15bp_MOTU06	1	HI28	33	0	67	0	100

15bp_MOTU05	10	D12_04 HI06 HI22 HI41 HI26 HI44 HI32 D12_03 HI10 HI23	43	56	0	1	100
15bp_MOTU04	6	CA47 CA45 CA41 CA22 CA02 CA40	100	0	0	0	100
15bp_MOTU03	18	HI09 HI03 HI42 HI18 HI34 HI07 HI15 HI37 HI35 HI02 HI04 HI21 HI12 HI33 HI16 HI30 HI11 CA14	100	0	0	0	100
15bp_MOTU02	21	HI39 HI40 HI29 HI08 HI27 HI43 HI36 HI20 HI31 HI01 FD01 HI17 HI46 D12_05 HI24 HI13 HI25 FD03 HI05 HI19 HI45	97	3	0	0	100
15bp_MOTU01	37	CA20 CA01 CA12 CA48 CA24 CA13 CA23 CA38 CA35 CA17 CA04 CA26 CA42 CA11 CA21 CA10 CA18 CA05 CA46 CA09 CA37 CA16 CA43 CA07 CA03 CA44 CA31 CA19 CA34 CA39 CA28 CA49 CA27 CA25 CA36 CA29 CA06	99	0	1	0	100
15bp	11		905	59	135	1	1100

Chapter 5
Discussion

Meiofaunal organisms present unique challenges for molecular barcoding due to unresolved questions of species definitions, species identification and genetic variation. The key issues for meiofaunal barcoding are the assumptions that protocols developed for large organisms will work for small organisms and a lack of standardisation across different surveys. Initial barcoding surveys sampled a broad range of taxa, where specimens were easily identified and readily available. As surveys have move towards intensive sampling of a few species across greater ranges, barcode data has highlighted distinct molecular taxa (MOTUs). Many of these MOTUs can be associated with different geographical locations, but not always. Sometimes, the molecular taxa are sympatric and can only be segregated by subtle ecological differences (e.g. Hebert *et al.*, 2004a). If barcodes had not been used to investigate the molecular taxonomy of the group, then we would have been none the wiser as to the complexity of the situation.

As well as being surprising, these types of cases raise a delicate issue within the traditional taxonomic community. How do we know that morphological designations are correct? In reference to larger animals, the answer is "probably". Although it requires time and occasionally specialist equipment, it is generally easy to identify whom mates with whom, what food species are consuming and where specimens are in their natural environment. With meiofauna, this is not the case. In comparison with megafauna, we know very little about meiofaunal organisms and due to their size, it is very difficult to study these organisms in their natural habitat. We don't always know whom mates with whom (if a species even requires sex for reproduction), what species are feeding on, or if a species is native to a particular site. Although it is possible to maintain some species as laboratory cultures, not all taxa will survive and the constant laboratory conditions will remove the natural variation of environmental abiotic factors. These factors alone cause problems within traditional meiofaunal morphological taxonomy when describing species. Adding to this situation that we know

only a small proportion of meiofaunal organisms have been described, and that there is a deficit of taxonomists, this means that it is unlikely that meiofauna can continue to be described based solely on morphology.
If meiofaunal identifications through barcoding are to be accepted across the global taxonomic community, there needs to be clearer communication between morphological and molecular taxonomists. There seems to have been misinterpretations of the undertakings of both groups, particularly circa 2003, when the number of barcoding surveys dramatically increased. During this period, there was no consistency between surveys as protocols, targets and analysis methods were being developed. Molecular barcoding has now been in routine use for almost a decade, and whilst there are standard protocols for sequence generation, a universal target and analysis method have yet to be adopted.

5.1 Is there a 'universal' target?

Initially, COI was proposed as a 'universal' target for barcoding (Hebert *et al.*, 2003a). Whilst this gene has been successfully used in a number of surveys, nematologists have shied away from COI, favouring instead the LSU or SSU genes for molecular surveys (Floyd *et al.*, 2002; Meldal *et al.*, 2007; Subbotin *et al.*, 2007; Ye *et al.*, 2007). Results from the investigations in this thesis suggest that, in some cases, LSU results may be used as a proxy measure for COI. Given the poor success rates of generating COI barcodes from meiofaunal taxa, the LSU should be targeted first. In addition, targeting the SSU will allow results from future surveys to be integrated with previous investigations, as there are more SSU sequences available in public databases. In October 2009, GenBank contained 7668 SSU entries compared with 3927 LSU and 1206 COI entries for the phylum Nematoda. For Tardigrada there were 753, 49 and 203 entries for SSU, LSU and COI respectively. Major barcoding initiatives may continue to insist on COI as the target, so attempts may be made to generate them. However, there may need to be a change in attitude regarding the virtues of COI.

5.2 *Is there a 'universal' cut-off?*

A value of 2% was designated as sufficient for species designation (Hebert *et al.*, 2003b). Again, this was based on results from large animals. Moreover this value was derived from analyses of the mitochondrial *cytB* gene, which shares the same mode of inheritance as COI, from a small survey of taxa (Johns and Avise, 1998). Although 2% may be sufficient for COI in some taxa, results from this thesis suggest that it is not suitable for LSU or SSU datasets. For some LSU data, a cut-off of 2% seems to be too high for species discrimination and results in clustering seen between genera (e.g. in the meiofaunal survey of Disko Island, Section 4.3.3). In other LSU data sets, 2% maybe considered as too low and is reporting individual variation (e.g. in the tardigrade survey, Section 4.3.4). These conflicting results are also seen in the SSU data.

Rather than assuming 2% is a suitable cut-off, it would be better practice to first generate the sequences, investigate the clustering behaviour, and then decide on a cut-off. Different surveys are likely to find different cut-offs which best describe the data. Ultimately, it is unlikely that there is a universal cut-off which can be used across multiple taxa. In studies where species identification is the key aim, this should not be an issue as the sequence data may define the cut-off. In bulk environmental surveys, where the diversity is being investigated, a cut-off may have to be middling value from the different taxa found within the sample, or different taxa need to be considered separately.

5.3 *Can taxa be defined by sequence similarity?*

Results from the thrips COI data (Chapter 2) demonstrated the ability of MOTU_define.pl and DOTUR to recover morphologically identified taxa at varying cut-offs. When MOTU_define.pl was challenged with anonymous sequences in Chapter 4, taxa were defined, some of which were supported by morphological identifications. Where barcodes are being used as a confirmational tool, the level of sequence similarity varies depending on the cut-off used and therefore so do the taxa defined. Molecular barcodes are

short DNA sequences which generally do not contain sufficient information to reconstruct phylogenetic processes. Using a taxon defining method based on an evolutionary model will overestimate taxa when sequences are of different lengths, as in DOTUR. Using a trimmed data set removes variation. If the evolutionary history of a set of samples is required, multiple, longer targets should be used. If molecular groups are sought, taxa can be adequately defined using simple comparisons of sequence similarity.

5.4 Can DNA barcodes identify meiofaunal taxa?

In meiofaunal taxa, it is unknown how morphological variation relates to molecular differences. Thus there is a tendency to perhaps overestimate the amount of morphological variation of meiofauna. The microscopic size of meiofauna means morphological identification is difficult and can lead to misidentification of specimens. DNA barcodes are a promising tool for the identification of any specimen, regardless of size. They can separate morphologically cryptic specimens and can reduce continuous morphological variation to robust molecular taxa. Where there is no morphological identification for a specimen, barcodes can provide some level of taxonomic assignment.

5.5 The future of meiofaunal barcoding

Previously, sequence generation had been limited to single individuals. It became possible to PCR from environmental bulk extracts but this required purification steps to remove PCR inhibitors and cloning in order to separate individual sequences. With the advent of next-generation sequencing platforms, there is the potential to barcode every individual within a sample; environmental metagenetic sequencing (Creer *et al.*, 2009). The amount of data generated by these surveys is vast, but are agnostic to any morphological information. Moreover, methods have not been optimised to enable relative abundances of individuals from a single taxon to be calculated from sequence frequencies (Porazinska *et al.*, 2009). Meiofaunal

organisms are a paraphyletic collection of taxa. Bulk extractions and sequencing may bias results through extraction methods, primer binding, DNA amplification and types of sequencing. In addition, accurately identifying species post sequencing requires a comprehensive database for comparisons (Machida *et al.*, 2009). Even so, it is normally possible to assign some level of taxonomic information to sequence data.

Ultrasequencing platforms will readily allow molecular variation to be recorded and assessed, but can make no assessment of the morphological diversity. Without similar input into measuring morphological diversity, there is likely to be a widening gap between the two. We will only be able to say what sequences exist in an environment, which may or may not be related to the biology of the sample. Integrating digital vouchering of specimens into surveys is possible and will enhance molecular results by expanding the range of taxa included in databases. Capturing morphology detail by VCE will allow meiofaunal diversity to be matched to molecular variation. There may be certain situations where VCE is not practical, but it should not be overlooked in the favour of molecular techniques and should be considered a priority for any sampling regime.

DNA taxonomy was proposed as an integration of molecular data and traditional taxonomy, when it became apparent that traditional taxonomy would be unable to complete the catalogue of life, especially the vast numbers of meiofauna taxa. However descriptions using DNA barcodes were felt to be lacking the rigorous and detailed study required by traditional taxonomy. Combining sequence generation with VCE would provide a permanent bridge between the demands of morphological and DNA taxonomy.

In its current state, meiofaunal barcoding is performing sub-optimally. Whilst exciting advances have been made, and the volumes of barcodes generated have increased exponentially, there are still gaps in the foundations. These need to be filled before we are able to generate an anonymous sequence and with certainty assign it to a single meiofaunal species and fully assess the meiofaunal diversity within a habitat.

References

Adams, B. J. (1998). "Species concepts and the evolutionary paradigm in modem nematology." J Nematol **30**(1): 1-21.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zang, Z. Zang, W. Miller and D. J. Lipman. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-3402.

Armstrong, K. F. and S. L. Ball (2005). "DNA barcodes for biosecurity: invasive species identification." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1813-1823.

Barucca, M., E. Olmo, S. Schiaparelli and A. Canapa (2004). "Molecular phylogeny of the family Pectinidae (Mollusca: Bivalvia) based on mitochondrial 16S and 12S rRNA genes." Mol Phylogenet Evol **31**(1): 89-95.

Bensasson, D., D. Zhang, D. L. Hartl and G. M. Hewitt (2001). "Mitochondrial pseudogenes: evolution's misplaced witnesses." Trends Ecol Evol **16**(6): 314-321.

Blaxter, M. (2003). "Molecular systematics: Counting angels with DNA." Nature **421**(6919): 122-124.

Blaxter, M., B. Elsworth, and J. Daub (2004). "DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades." Proc Biol Sci **271 Suppl 4**: S189-192.

Blaxter, M. and R. Floyd (2003). "Molecular taxonomics for biodiversity surveys: already a reality." Trends in Ecology & Evolution **18**(6): 268-269.

Blaxter, M., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd and E. Abebe (2005). "Defining operational taxonomic units using DNA barcode data." Philos Trans R Soc Lond B Biol Sci **360**(1462): 1935-1943.

Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman, A. Vierstraete, J. R. Vanfleteren, L. Y. Mackey, M. Dorris, L. M. Frisse, J. T. Vida and W. K. Thomas (1998). "A molecular evolutionary framework for the phylum Nematoda." Nature **392**(6671): 71-75.

Bongers, T. (1994). De Nematoden van Nederland. Utrecht.

Brunner, P. C., E. K. Chatzivassiliou, N. I. Katis and J. E. Frey (2004). "Host-associated genetic differentiation in Thrips tabaci (Insecta; Thysanoptera), as determined from mtDNA sequence data." Heredity **93**(4): 364-370.

Brunner, P. C., C. C. Fleming and J. E. Frey (2002). "A molecular identification key for economically important thrips species (Thysanoptera: Thripidae) using direct sequencing and a PCR-RFLP-based approach." Agricultural and Forest Entomology **4**: 127-136.

Cognato, A. I. (2006). "Standard percent DNA sequence difference for insects does not predict species boundaries." Journal of Economic Entomology **99**(4): 1037-1045.

Creer, S., V. Fonseca, D. L. Porazinska, R. M. Giblin-Davis, W. Sung, D. M. Power, M. Packer, G. R. Carvalho, M. Blaxter, P. J. D. Lambshead, and W. K. Thomas. (2009). "Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises." Molecular Ecology ***In press***.

Crespi, B. J., D. Carmean, L. Vawter and C. Dohlen (1996). "The molecular systematics of Thysanoptera." Sytematic Entomology **21**: 79-87.

Crespi, B. J., D. A. Carmean, L. A. Mound, M. Worobey and D. Morris (1998). "Phylogenetics of social behavior in Australian gall-forming thrips: Evidence from mitochondrial DNA sequence, adult morphology and behavior, and gall morphology." Molecular Phylogenetics and Evolution **9**(1): 163-180.

Dawson, M. N., K. A. Raskoff and D. K. Jacobs (1998). "Field preservation of marine invertebrate tissue for DNA analyses." Molecular Marine Biology and Biotechnology **7**(2): 145-152.

De Ley, P. and W. Bert (2002). "Video Capture and Editing as a Tool for the Storage, Distribution, and Illustration of Morphological Characters of Nematodes." Journal of Nematology **34**(4): 296-302.

De Ley, P. and M. Blaxter (2002). Systematic Position and Phylogeny. The Biology of Nematodes. D. L. Lee. London, Taylor & Francis**:** 1-30.

De Ley, P., I. T. De Ley, K. Morris, E. Abebe, M. Mundo-Ocampo, M. Yoder, J. Heras, D. Waumann, A. Rocha-Olivares, A. H. Jay Burr, J. G. Baldwin and W. K. Thomas (2005). "An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding." Philos Trans R Soc Lond B Biol Sci **360**(1462): 1945-1958.

Derycke, S., T. Remerie, T. Remerie, A. Vierstraete, T. Backeljau, J. R. Vanfleteren, M. Vincx and T. Moens (2005). "Mitochondrial DNA variation and cryptic speciation within the free-living marine nematode *Pellioditis marina*." Marine Ecology Progress Series **300**: 91-103.

DeSalle, R., M. G. Egan and M. Siddall (2005). "The unholy trinity: taxonomy, species delimitation and DNA barcoding." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1905-1916.

Dorris, M., M. E. Viney and M. Blaxter (2002). "Molecular phylogenetic analysis of the genus *Strongyloides* and related nematodes." International Journal for Parasitology **32**: 1507-1517.

Dunn, C. P. (2003). "Keeping taxonomy based in morphology." Trends in Ecology & Evolution **18**(6): 270-271.

Ebach, M. C. and C. Holdrege (2005). "DNA barcoding is no substitute for taxonomy." Nature **434**(7034): 697-697.

Elias, M., R. I. Hill, K. R. Willmott, K. K. Dasmahapatra, A. V. Z. Brower, J. Mallet and C. D. Jiggins (2007). "Limited performance of DNA barcoding in a diverse community of tropical butterflies." Proceedings of the Royal Society B-Biological Sciences **274**(1627): 2881-2889.

Evans, K. M., A. H. Wortley and D. G. Mann (2007). "An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta)." Protist **158**(3): 349-364.

Floyd, R., E. Abebe, A. Papert and M. Blaxter (2002). "Molecular barcodes for soil nematode identification." Mol Ecol **11**(4): 839-850.

Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek (1994). "DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates." Molecular Marine Biology and Biotechnology **3**(5): 294-299.

Fouquet, A., A. Gilles, M. Vences, C. Marty, M. Blanc and N. J. Gemmell (2007). "Underestimation of species richness in neotropical frogs revealed by mtDNA analyses." PLoS ONE **2**(10): e1109.

Gasser, R. B., N. B. Chilton, H. Hoste and L. A. Stevenson (1994). "Species identification of trichostrongyle nematodes by PCR-linked RFLP." Int J Parasitol **24**(2): 291-293.

Gewin, V. (2002). "Taxonomy - All living things, online." Nature **418**(6896): 362-363.

Godfray, H. C. J. (2002). "Challenges for taxonomy - The discipline will have to reinvent itself if it is to survive and flourish." Nature **417**(6884): 17-19.

Gregory, T. R. (2005). "DNA barcoding does not compete with taxonomy." Nature **434**(7037): 1067-1067.

Gullan, P. J. and P. S. Cranston (2005). The Insects: An Outline of Entomology, Blackwell Publishing Ltd.

Hajibabaei, M., J. R. deWaard, N. V. Ivanova, S. Ratnasingham, R. T. Dooh, S. L. Kirk, P. M. Mackie and P. D. Hebert (2005). "Critical factors for assembling a high volume of DNA barcodes." Philos Trans R Soc Lond B Biol Sci **360**(1462): 1959-1967.

Hajibabaei, M., D. H. Janzen, J. M. Burns, W. Hallwachs and P. D. Hebert (2006). "DNA barcodes distinguish species of tropical Lepidoptera." Proc Natl Acad Sci U S A **103**(4): 968-971.

Hebert, P. D., A. Cywinska, S. L. Cywinska S. L. Ball and J. R. deWaard (2003a). "Biological identifications through DNA barcodes." Proc Biol Sci **270**(1512): 313-321.

Hebert, P. D., S. Ratnasingham and J. R. deWaard (2003b). "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species." Proc Biol Sci **270 Suppl 1**: S96-99.

Hebert, P. D., E. H. Penton, J. M. Burns, D. H. Janzen and W. Hallwachs (2004a). "Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*." Proc Natl Acad Sci U S A **101**(41): 14812-14817.

Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak and C. M. Francis (2004b). "Identification of birds through DNA barcodes." Plos Biology **2**(10): 1657-1663.

Hellborg, L., I. Gunduz and M. Jaarola (2005). "Analysis of sex-linked sequences supports a new mammal species in Europe." Mol Ecol **14**(7): 2025-2031.

Hingley, M. (1993). Microscopic life in *Sphagnum*. Slough, The Richmond Publishing Co. Ltd.

Hosseini, R., M. A. Keller, O. Schmidt and V. W. Framenau (2007). "Molecular identification of wolf spiders (Araneae : Lycosidae) by multiplex polymerase chain reaction." Biological Control **40**(1): 128-135.

Huber, J. A., D. Mark Welch, H. G. Morrison, S. M. Huse, P. R. Neal, D. A. Butterfield and M. L. Sogin (2007). "Microbial population structures in the deep marine biosphere." Science **318**(5847): 97-100.

Hunt, D. J. (1993). Aphelenchida, Longidoridae and Trichodoridae: Their Systematics and Bionomics. Cambridge, CAB International.

Hurst, G. D. and F. M. Jiggins (2005). "Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts." Proc Biol Sci **272**(1572): 1525-1534.

Inoue, T. and T. Sakurai (2007). "The phylogeny of thrips (Thysanoptera : Thripidae) based on partial sequences of cytochrome oxidase I, 28S ribosomal DNA and elongation factor-1 alpha and the association with vector competence of tospoviruses." Applied Entomology and Zoology **42**(1): 71-81.

Ivanova, N. V., A. V. Borisenko and P. D. Hebert (2009). "Express barcodes: racing from specimen to identification." Molecular Ecology Resources **9**(Suppl. 1): 35-41.

Johns, G. C. and J. C. Avise (1998). "A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene." Molecular Biology and Evolution **15**(11): 1481-1490.

Lawton, J. H., D. E. Bignell, D. E. Bignell, B. Bolton, G. F. Bloemers, P. Eggleton, P. M Hammond, M. Hodda, R. D. Holt, T. B. Larsen, N. A. Mawdsley, N. E. Stork, D. S. Srivastava and A. D. Watt (1998). "Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest." Nature **391**(6662): 72-76.

Lee, M. S. Y. (2000). "A worrying systematic decline." Trends in Ecology & Evolution **15**(8): 346-346.

Lipscomb, D., N. Platnick and Q. Wheeler (2003). "The intellectual content of taxonomy: a comment on DNA taxonomy." **18**(2): 65-66.

Lohse, K. (2009). "Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006)." Systematic Biology **58**(4): 439-441.

Lorenz, J. G., W. E. Jackson, J. C. Beck and R. Hanner (2005). "The problems and promise of DNA barcodes for species diagnosis of primate biomaterials." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1869-1877.

Machida, R. J., Y. Hashiguchi, M. Nishida and S. Nishida (2009). "Zooplankton diversity analysis through single-gene sequencing of a community sample." BMC Genomics **10**: 438-444.

Mayr, E. (1942). "Systematics and the origins of species from the viewpoint of a zoologist." New York: Columbia University Press.

McLeish, M. J., T. W. Chapman and L. A. Mound (2006). "Gall morpho-type corresponds to separate species of gall-inducing thrips (Thysanoptera : Phlaeothripidae)." Biological Journal of the Linnean Society **88**(4): 555-563.

Meier, R., K. Shiyang, G. Vaidya and P. K. L. Ng (2006). "DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success." Syst Biol **55**(5): 715-728.

Meldal, B. H. M., N. J. Debenham, P. De Ley, I. T. De Ley, J. R. Vanfleteren, A. R. Vierstraete, W. Bert, G. Borgonie, T. Moens, P. A. Tyler, M. C. Austen, M. L. Blaxter, A. D. Rogers and P. J. D. Lambshead (2007). "An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa." Molecular Phylogenetics and Evolution **42**(3): 622-636.

Meyer, C. P. and G. Paulay (2005). "DNA barcoding: error rates based on comprehensive sampling." PLoS Biol **3**(12): e422.

Minelli, A. (2003). "The status of taxonomic literature." Trends in Ecology & Evolution **18**(2): 75-76.

Monaghan, M. T., M. Balke, T. R. Gregory and A. P. Vogler (2005). "DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers." Philos Trans R Soc Lond B Biol Sci **360**(1462): 1925-1933.

Morgan, C. I. and P. E. King (1976). British Tardigrades. London, The Linnean Society of London.

Moritz, C. and C. Cicero (2004). "DNA barcoding: Promise and pitfalls." Plos Biology **2**(10): 1529-1531.

Mound, L. A., B. S. Heming and J. M. Palmer (1980). "Phylogenetic relationships between the families of recent Thysanoptera (Insecta)." Zoological Journal of the Linnean Society **69**: 111-141.

Mound, L. A. and D. C. Morris (2001). "Domicile constructing phlaeothripine Thysanoptera from Acacia phyllodes in Australia: Dunatothrips Moulton and Sartrithrips gen.n., with a key to associated genera." Systematic Entomology **26**(4): 401-419.

Murrell, A. and S. C. Barker (2005). "Multiple origins of parasitism in lice: phylogenetic analysis of SSU rDNA indicates that the Phthiraptera and Psocoptera are not monophyletic." Parasitology Research **97**(4): 274-280.

Nadler, S. A. (2002). "Species delimitation and nematode biodiversity: phylogenies rule." Nematology **4**: 615-625.

Nadler, S. A., E. Bolotin and S. P. Stock (2006). "Phylogenetic relationships of Steinernema Travassos, 1927 (Nematoda: Cephalobina: Steinernematidae) based on nuclear, mitochondrial and morphological data." Syst Parasitol **63**(3): 161-181.

Packer, L., J. Gibbs, et al. (2009). "DNA barcoding and the mediocrity of morphology." Molecular Ecology Resources **9**: 42-50.

Paquin, P. and M. Hedin (2004). "The power and perils of 'molecular taxonomy': a case study of eyeless and endangered Cicurina (Araneae : Dictynidae) from Texas caves." Molecular Ecology **13**(10): 3239-3255.

Petersen, B. (1951). The Tardigrade fauna of Greenland.

Pfenninger, M., C. Nowak, C. Kley, D. Steinke and B. Streit (2007). "Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic Chironomus (Diptera) species." Mol Ecol **16**(9): 1957-1968.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin and A. P. Vogler (2006). "Sequence-based species delimitation for the DNA taxonomy of undescribed insects." Syst Biol **55**(4): 595-609.

Porazinska, D. L., R. M. Giblin-Davis, L. Faller, W. Farmerie, N. Kanzaki, K. Morris, T. Powers, A. Tucker, W. Sung and W. K. Thomas (2009). "Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity." Molecular Ecology Resources **9**: 1439-1450.

Quince, C., T. P. Curtis and W. T. Sloan (2008). "The rational exploration of microbial diversity." Isme Journal **2**(10): 997-1006.

Ratnasingham, S. and P. D. Hebert (2007). "bold: The Barcode of Life Data System (http://www.barcodinglife.org)." Mol Ecol Notes **7**(3): 355-364.

Read, D. S., S. K. Sheppard, M. W. Bruford, D. M. Glen and W. O. C. Symondson (2006). "Molecular detection of predation by soil micro-arthropods on nematodes." Molecular Ecology **15**(7): 1963-1972.

Riggin, G. T. J. (1962). Tardigrada of Southwest Virginia: With the addition of a description of a new marine species from Florida.

Roe, A. D. and F. A. H. Sperling (2007). "Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding." Molecular Phylogenetics and Evolution **44**(1): 325-345.

Ronquist, F. and U. Gardenfors (2003). "Taxonomy and biodiversity inventories: time to deliver." Trends Ecol Evol **18**(6): 269-270.

Ros, V. I. D. and J. A. J. Breeuwer (2007). "Spider mite (Acari : Tetranychidae) mitochondrial COI phylogeny reviewed: host plant relationships, phylogeography, reproductive parasites and barcoding." Experimental and Applied Acarology **42**(4): 239-262.

Rubinoff, D. (2006). "Utility of mitochondrial DNA barcodes in species conservation." Conservation Biology **20**(4): 1026-1033.

Schloss, P. D. and J. Handelsman (2005). "Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness." Appl Environ Microbiol **71**(3): 1501-1506.

Seberg, O., C. J. Humphries, S. Knapp, D. W. Stevenson, G. Petersen, N. Scharff and N. M. Andersen (2003). "Shortcuts in systematics? A commentary on DNA-based taxonomy." Trends in Ecology & Evolution **18**(2): 63-65.

Seutin, G., B. N. White and P. T. Boag (1991). "Preservation of avian blood and tissue samples for DNA analyses." Can. J. Zool. **69**: 82-90.

Slapeta, J., P. Lopez-Garcia and D. Moreira (2006). "Global dispersal and ancient cryptic species in the smallest marine eukaryotes." Molecular Biology and Evolution **23**(1): 23-29.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta and G. J. Herndl (2006). "Microbial diversity in the deep sea and the underexplored "rare biosphere"." Proceedings of the National Academy of Sciences of the United States of America **103**(32): 12115-12120.

Srivastava, M., E. Begovic, J. Chapman, N. H. Putnam, U. Hellsten, T. Kawashima, A. Kuo, T. Mitros, A. Salamov, M. L. Carpenter, A. Y. Signorovitch, M. A. Moreno, K. Kamm, J. Grimwood, J. Schmutz, H. Shapiro, I. V. Grigoriev, L. W. Buss, B. Schierwater, S. L. Dellaporta and D. S. Rokhsar (2008). "The Trichoplax genome and the nature of placozoans." Nature **454**(7207): 955-U919.

Subbotin, S. A., D. Sturhan, N. Vovlas, P. Castillo, J. T. Tambe, M. Moens and J. G. Baldwin (2007). "Application of the secondary structure model of rRNA for phylogeny: D2-D3 expansion segments of the LSU gene of plant-parasitic nematodes from the family Hoplolaimidae Filipjev, 1934." Molecular Phylogenetics and Evolution **43**(3): 881-890.

Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), Sinauer Associates, Sunderland, Massachusetts.

Tautz, D., P. Arctander, A. Minelli, R. H. Thomas and A. P. Vogler (2002). "DNA points the way ahead of taxonomy - In assessing new approaches, it's time for DNA's unique contribution to take a central role." Nature **418**(6897): 479-479.

Tautz, D., P. Arctander, M. Alessandro, R. H. Thomas and A. P. Vogler (2003). "A plea for DNA taxonomy." Trends Ecol Evol **18**(2): 70-74.

Toda, S. and T. Murai (2007). "Phylogenetic analysis based on mitochondrial COI gene sequences in Thrips tabaci Lindeman (Thysanoptera : Thripidae) in relation to reproductive forms and geographic distribution." Applied Entomology and Zoology **42**(2): 309-316.

Trewick, S. A. (2008). "DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera : Acrididae)." Cladistics **24**(2): 240-254.

Vences, M., M. Thomas, R. M. Bonett and D. R. Vieites (2005). "Deciphering amphibian diversity through DNA barcoding: chances and challenges." Philosophical Transactions of the Royal Society B-Biological Sciences **360**(1462): 1859-1868.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Y. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers and H. O. Smith (2004). " Environmental genome shotgun sequencing of the Sargasso Sea." Science **304**(5667): 66-74.

Waugh, J. (2007). "DNA barcoding in animal species: progress, potential and pitfalls." BioEssays **29**: 188-197.

Wiemers, M. and K. Fiedler (2007). "Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae)." Frontiers in Zoology **4**(8).

Wilson, E. O. (2003). "The encyclopedia of life." Trends Ecol Evol **18**(2): 77-80.

Witt, J. D., D. L. Threloff and P. D. Hebert (2006). "DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation." Mol Ecol **15**(10): 3073-3082.

Ye, W., R. M. Giblin-Davis, H. Braasch, K. Morris and W. K. Thomas (2007). "Phylogenetic relationships among Bursaphelenchus species (Nematoda: Parasitaphelenchidae) inferred from nuclear ribosomal and mitochondrial DNA sequence data." Mol Phylogenet Evol **43**(3): 1185-1197.

Yoder, M., I. Tandingan De Ley, I. W. King, M. Mundo-Ocampo, J. Mann, M. Blaxter, L. Poiras and P. De Ley (2006). "DESS: a versatile solution for preserving morphology and extractable DNA of nematodes." Nematology **8**(3): 367-376.

Yoshizawa, K. and K. P. Johnson (2005). "Aligned 18S for Zoraptera (Insecta): Phylogenetic position and molecular evolution." Molecular Phylogenetics and Evolution **37**(2): 572-580.